

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number  
**WO 01/98351 A2**

(51) International Patent Classification<sup>7</sup>: C07K 14/705

(21) International Application Number: PCT/US01/19275

(22) International Filing Date: 15 June 2001 (15.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/212,483	16 June 2000 (16.06.2000)	US
60/213,954	22 June 2000 (22.06.2000)	US
60/215,209	29 June 2000 (29.06.2000)	US
60/216,595	7 July 2000 (07.07.2000)	US
60/218,936	14 July 2000 (14.07.2000)	US
60/219,154	19 July 2000 (19.07.2000)	US
60/220,141	21 July 2000 (21.07.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). NGUYEN, Dannel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). KALLICK, Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). KHAN, Farrah, A. [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA

94303 (US). WALLA, Narinder, K. [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). GRAUL, Richard [US/US]; 682-29th Avenue, San Francisco, CA 94121 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). HERNANDEZ, Roberto [GB/GB]; 160 Pine Tree Avenue, Canterbury Kent, CT2 7SZ (GB). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). HE, Ann [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.



WO 01/98351 A2

## G-PROTEIN COUPLED RECEPTORS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of G-protein coupled  
5 receptors and to the use of these sequences in the diagnosis, treatment, and prevention of cell  
proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic  
disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the  
expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

### BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals.  
Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a  
hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus  
activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular  
15 target molecule, such as a transcription factor. This process of signal transduction regulates all types  
of cell functions including cell proliferation, differentiation, and gene transcription. The G-protein  
coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a  
central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a  
proven history of being successful therapeutic targets.

20 GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic  
transmembrane domains which together form a bundle of antiparallel alpha ( $\alpha$ ) helices. GPCRs range  
in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10;  
Coughlin, S.R. (1994) Curr. Opin. Cell Biol. 6:191-197). The amino-terminus of a GPCR is  
extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic  
25 and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the  
transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops  
may interact with agonists and antagonists. The most conserved domains of GPCRs are the  
transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account,  
in part, for structural and functional features of the receptor. In most cases, the bundle of  $\alpha$  helices  
30 forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three  
extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by  
inducing a conformational change in intracellular portions of the receptor. In turn, the large, third  
intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding  
(G) protein complex which mediates further intracellular signaling activities, including the activation  
35 of second messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and

the interaction of the activated GPCR with ion channel proteins. (See, .g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J.M. (1994) *Curr. Opin. Cell Biol.* 6:180-190.)

5 GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine,  $\gamma$ -aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid  
10 mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

15 The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are less common. Some receptors have more than one site at which variance  
20 can occur. The splicing variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) *Trends Pharmacol. Sci.* 20:294-301).

GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar  
25 functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by  
30 triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and glutamate (N-methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall  
35 (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22,

32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783.)

The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and Alzheimer's disease (Kask, K. et al. (1997) Life Sci. 60:1523-1533). Other nervous system rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al. (1999) Cell Biochem. Biophys. 30:213-242).

The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-like GPCR family. These receptors function by transducing odorant signals. Numerous distinct olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors are found in nasal passages. For example, the RA1c receptor which was isolated from a rat brain library, has been shown to be limited in expression to very distinct regions of the brain and a defined zone of the olfactory epithelium (Raming, K. et al. (1998) Receptors Channels 6:141-151). However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example, three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed expression patterns not only in taste and olfactory tissue, but also in male reproductive tissue (Thomas, M.B. et al. (1996) Gene 178:1-5).

Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine (Watson, *supra*, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the production of cAMP.

Examples of secretin-like GPCRs implicated in inflammation and the immune response include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily. These seven transmembrane hormone receptors exist as heterodimers *in vivo* and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes and is markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon (1998) J. Leukoc. Biol. 63:271-280).

The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the

major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, supra, p.130). The  $\text{Ca}^{2+}$ -sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the  $\text{GABA}_B$  receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold Dictyostelium discoideum, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, supra). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin  $\text{V}_2$  (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcemia, hypercalcemia); parathyroid hormone (short limbed dwarfism);  $\beta_3$ -adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmacol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, supra; Stadel, supra). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and

tachycardia; serotonin 5HT1D antagonists are used against migraine; and histamine H1 antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, supra).

Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V2 vasopressin receptors causing nephrogenic diabetes could be functionally rescued in vitro by co-expression of a C-terminal V2 receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291). Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V2 receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R.G. MacKenzie (1999) J. Biol. Chem. 274:35816-35822), and thus might be treated with a similar strategy. The type 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Mannstadt, M. et al. (1999) Am. J. Physiol. 277:F665-F675).

The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) Annu. Rev. Med. 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1, results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

The discovery of new G-protein coupled receptors and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

#### SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to

collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," and "GCREC-17." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-17.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-17. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:18-34.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.



The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test

compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:18-34, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including

predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Table 8 shows tissue-specific expression of polynucleotides of the invention.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and

human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with  
5 GCREC or by acting on components of the biological pathway in which GCREC participates.

An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to  
10 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or  
15 a polypeptide with at least one functional characteristic of GCREC. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be "altered," and may contain deletions, insertions,  
20 or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged  
25 amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide,  
30 polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.  
35 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well

known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by  
5 directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using  
10 fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize  
15 the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures  
20 on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as  
25 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring  
30 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"  
35 refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide

thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,

5 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may  
10 be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated  
15 DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and  
20 assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded  
25 as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
30	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
35	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile

	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
5	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

---

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of GCREC or the polynucleotide encoding GCREC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.



A fragment of SEQ ID NO:18-34 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:18-34, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:18-34 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish  
5 SEQ ID NO:18-34 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:18-34 and the region of SEQ ID NO:18-34 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-17 is encoded by a fragment of SEQ ID NO:18-34. A fragment of SEQ ID NO:1-17 comprises a region of unique amino acid sequence that specifically identifies  
10 SEQ ID NO:1-17. For example, a fragment of SEQ ID NO:1-17 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-17. The precise length of a fragment of SEQ ID NO:1-17 and the region of SEQ ID NO:1-17 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

15 A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

20 The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

25 Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS  
30 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms  
35 is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment

Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

15 *Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

20 Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

25 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

30 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the

site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized

after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill  
5 in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature  
10 under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al.  
15 (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC  
20 concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions  
25 will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A  
30 hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed):

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide  
35 sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

5 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

10 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of GCREC. For example, modulation  
15 may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GCREC.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the  
20 antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where  
25 necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript  
30 elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of GCREC.

35 "Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments

thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule.

Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

“Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping

Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing GCRC, nucleic acids encoding GCRC, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or

cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in



vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection,

5 transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of  
10 the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an  
15 "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to  
20 another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a  
25 propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at  
30 least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

35 The invention is based on the discovery of new human G-protein coupled receptors

(GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:1 is 99% identical to human orphan G protein-coupled receptor; GPC-R (GenBank ID g2865470) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-217, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains an orphan G protein-coupled receptor domain as determined by searching for statistically significant matches in the DMO database using a Blocks IMProved Searcher (BLIMPS) that searches for gene families, sequence

homology, and structural fingerprint regions. (See Table 3.) SEQ ID NO:1 additionally contains a G-protein coupled receptor subclass 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS-PRINTS, BLIMPS-BLOCKS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:8 is 79% identical to rat serotonin receptor (GenBank ID g310075) as determined by BLAST. (See Table 2.) The BLAST probability score is  $2.0e-151$ . SEQ ID NO:8 also contains G-protein coupled receptor domain structure as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses, as well as BLAST comparisons to protein signature sequences in the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:8 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:10 is 92% identical to the human leukocyte platelet-activating factor receptor, a G-protein coupled receptor (GenBank ID g189270), as determined by BLAST. (See Table 2.) The BLAST probability score is  $2.9e-147$ . SEQ ID NO:10 also contains 7-transmembrane receptor domains, characteristic of G-protein coupled receptors, as well as a receptor binding domain, as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:10 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:11 is 99% identical to human endothelin receptor B delta 3 (GenBank ID g4580924), as determined by BLAST. (See Table 2.) The BLAST probability score is  $2.3e-289$ . SEQ ID NO:11 also contains 7-transmembrane receptor domains as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is an endothelin receptor. In an alternative example, SEQ ID NO:16 is 50% identical to the rat taste bud receptor protein TB641 (GenBank ID g1256393), as determined by BLAST. (See Table 2.) The BLAST probability score is  $2.0e-59$ . BLAST searches of the PRODOM and DOMO databases also indicate that SEQ ID NO:16 has homology with receptor protein and G-protein coupled receptor domains. SEQ ID NO:16 also contains a rhodopsin family 7-transmembrane receptor domain, as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a G-protein coupled receptor. In an alternative example, SEQ ID NO:17 is 41% identical to the rat G-protein coupled receptor OL1 (GenBank ID g1016362), as determined by BLAST. (See Table 2.) The BLAST probability score is

3.6e-69. SEQ ID NO:17 also contains a 7-transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is a G-protein coupled receptor of the rhodopsin family. SEQ ID NO:2-7, SEQ ID NO:9, and SEQ ID NO:12-15 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-17 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:18-34 or that distinguish between SEQ ID NO:18-34 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2432516H1 is the identification number of an Incyte cDNA sequence, and BRAVUNT02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71687857V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6479070.edit is the identification number of a Genscan-predicted coding sequence, with g6479070 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL210313\_00001 represents a "stitched" sequence in which 210313 is the identification number of the cluster of sequences to

which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL7655614\_g7406476\_g156725 is the identification number of a  
5 "stretched" sequence, with 7655614 being the Incyte project identification number, g7406476 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g156725 being the GenBank identification number of the nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide  
10 sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors  
15 which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows tissue-specific expression of polynucleotides of the invention. Column 1 lists groups of tissues which were tested by polymerase chain reaction (PCR) for expression of the polynucleotides. The remaining columns indicate whether a particular polynucleotide was expressed in each tissue group. Detection of a PCR product indicated positive expression, denoted by a "+"  
20 sign, while inability to detect a PCR product indicated a lack of expression, denoted by a "-" sign.

The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

25 The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:18-34, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the  
30 sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a  
35 polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID

NO:18-34 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:18-34. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These  
10 combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately  
15 selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the  
20 nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the  
25 synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID  
30 NO:18-34 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of  
35 the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment

of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding GCREC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GCREC.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random



point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7*:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7*:225-232.)

Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used.

(See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCREC. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCREC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCREC can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCREC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol.*

Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GCREC. A number of vectors  
5 containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994)  
10 *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of GCREC. Transcription of sequences encoding GCREC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock  
15 promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

20 In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc.*  
25 *Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are  
30 constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed  
35 into cell lines using expression vectors which may contain viral origins of replication and/or

endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which  
5 successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> and *apv*<sup>-</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic,  
10 or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981)  
15 J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to  
20 quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing  
25 sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that  
30 express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC using either  
35 specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques

include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, 5 e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and 10 may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, 15 and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic 20 agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors 25 containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, 30 phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct 35 modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GCREC may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of GCREC, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or

E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is  
5 detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical  
10 libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GCREC of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of GCREC. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for  
15 GCREC activity, wherein GCREC is combined with at least one test compound, and the activity of GCREC in the presence of a test compound is compared with the activity of GCREC in the absence of the test compound. A change in the activity of GCREC in the presence of the test compound is indicative of a compound that modulates the activity of GCREC. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising GCREC under conditions suitable for  
20 GCREC activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of GCREC may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic  
25 stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo;  
30 Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic  
Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell  
35 blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred

to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding GCREC may also be manipulated in vitro in ES cells derived from  
5 human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding GCREC can also be used to create "knockin" humanized animals  
10 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCREC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease.  
15 Alternatively, a mammal inbred to overexpress GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists  
20 between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with breast tissue, megakaryoblast cells, prostate tumor, dorsal root ganglion tissue, and pituitary gland tissue. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or  
25 activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or  
30 activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of  
35 the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,



gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders,

5 amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal

10 familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system

15 disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a

20 cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular

25 heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such

30 as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the

liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, 5 Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an

10 autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, 15 erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome,

20 systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, 25 hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus.

In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

30 In a further embodiment, a composition comprising a substantially purified GCREC in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GCREC may be

administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those listed above.

In a further embodiment, an antagonist of GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, described above. In one aspect, an antibody which specifically binds GCREC may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GCREC.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will

consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

- 5 Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and  
10 Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

- In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda,  
15 S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g.,  
Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

- 20 Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

- Antibody fragments which contain specific binding sites for GCREC may also be generated.  
25 For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

- 30 Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GCREC and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies  
35 reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay

may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding GCREC, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GCREC. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,

Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other  
5 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for  
10 somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475),  
15 cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated  
20 cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the  
25 case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of GCREC from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing  
30 these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr.*  
35 *Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of GCRC include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCRC may be expressed using (i) a  
5 constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the  
10 ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GCRC from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID  
15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of  
20 these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCRC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCRC under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive  
25 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for  
30 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant")  
35 discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by

reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; 5 Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well 10 known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) 15 Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be 20 especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. 25 Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 30 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of 35 ordinary skill in the art.



In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During  
5 alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the  
10 alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will  
15 allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

20 Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using  
25 triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of  
30 RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by  
35 scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,

GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding GCRC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCRC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCRC expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding GCRC may be therapeutically useful, and in the treatment of disorders associated with decreased GCRC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCRC may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCREC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins.

Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

5           The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

          Compositions for pulmonary administration may be prepared in liquid or dry powder form.  
10       These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton,  
15       J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

          Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

20       Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to  
25       transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

          For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration  
30       range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

          A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be  
35       determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such

as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are  
5 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the  
10 subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week,  
15 or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their  
20 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being  
25 treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter  
30 molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell  
35 extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to

GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for  
5 diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GCREC may be correlated  
10 with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is  
15 made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding GCREC, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%  
20 sequence identity to any of the GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:18-34 or from genomic sequences including promoters, enhancers, and introns of the GCREC gene.

Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC derivatives into vectors for the  
25 production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

30 Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders associated with expression of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including  
35 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other

5 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous

10 system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial

15 nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic

20 neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure,

25 ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart

30 disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis,

pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus. The polynucleotide sequences encoding GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a



suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to  
5 amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-  
10 throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence  
15 chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Dupl  , C.  
20 et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the  
25 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor  
30 progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her  
35 pharmacogenomic profile.

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

5 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number  
10 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The  
15 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present  
20 invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson  
25 (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is  
30 not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to  
35 prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of

Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem.

270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

5 Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which  
10 alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated  
15 biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the  
20 polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological  
25 sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci.*  
30 *USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

35 In another embodiment of the invention, nucleic acid sequences encoding GCRC may be

used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

between GCREC and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

10 In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCREC specifically compete with a test compound for binding GCREC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCREC.

15 In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

20 The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/212,483, U.S. Ser. No. 60/213,954, U.S. Ser. No. 60/215,209, U.S. Ser. No. 60/216,595, U.S. Ser. No. 60/218,936, U.S. Ser. No. 60/219,154, and U.S. Ser. No. 60/220,141, are expressly incorporated by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

25 Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).



### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal  
5 cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides  
10 were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques  
15 disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public  
20 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The  
25 Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on  
30 GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite,  
35 and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length

polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program  
5 (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate  
10 references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ  
15 ID NO:18-34. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative G-protein coupled receptors were initially identified by running the Genscan gene  
20 identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of  
25 Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been  
30 annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted  
35 sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available,

this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

##### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The

GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

5 **VI. Chromosomal Mapping of GCREC Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:18-34 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:18-34 were assembled into clusters of contiguous and overlapping sequences using  
10 assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

15 Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM  
20 distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

25 **VII. Analysis of Polynucleotide Expression**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

30 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

35

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding GCREC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding GCREC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of GCREC Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was

synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was  
5 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

10 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:18-34 are employed to screen cDNAs,  
15 genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase  
20 (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

25 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and  
30 compared.

#### **X. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the  
35 aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999),

supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation



Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  
5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR  
10 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average  
15 concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate  
20 buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### **Hybridization**

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample  
25 mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash  
30 buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines

at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### **XI. Complementary Polynucleotides**

Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same

procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GCREC-encoding transcript.

## XII. Expression of GCREC

Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-

thiogalactopyranoside (IPTG). Expression of GCREC in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GCREC by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays

shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

### XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-GCREC activity by, for example, binding the peptide or GCREC to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring GCREC Using Specific Antibodies**

Naturally occurring or recombinant GCREC is substantially purified by immunoaffinity chromatography using antibodies specific for GCREC. An immunoaffinity column is constructed by covalently coupling anti-GCREC antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GCREC are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GCREC (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GCREC binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GCREC is collected.

#### **XVI. Identification of Molecules Which Interact with GCREC**

Molecules which interact with GCREC may include agonists and antagonists, as well as molecules involved in signal transduction, such as G proteins. GCREC, or a fragment thereof, is labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) A fragment of GCREC includes, for example, a fragment comprising one or more of the three extracellular loops, the extracellular N-terminal region, or the third intracellular loop. Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed, and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate ligand molecules.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid

system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
5 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential GCREC agonists or antagonists may be tested for activation or inhibition of GCREC receptor activity using the assays described in sections XVII and XVIII. Candidate molecules may be selected from known GPCR agonists or antagonists, peptide libraries, or  
10 combinatorial chemical libraries.

Methods for detecting interactions of GCREC with intracellular signal transduction molecules such as G proteins are based on the premise that internal segments or cytoplasmic domains from an orphan G protein-coupled seven transmembrane receptor may be exchanged with the analogous domains of a known G protein-coupled seven transmembrane receptor and used to identify  
15 the G-proteins and downstream signaling pathways activated by the orphan receptor domains (Kobilka, B.K. et al. (1988) Science 240:1310-1316). In an analogous fashion, domains of the orphan receptor may be cloned as a portion of a fusion protein and used in binding assays to demonstrate interactions with specific G proteins. Studies have shown that the third intracellular loop of G protein-coupled seven transmembrane receptors is important for G protein interaction and signal  
20 transduction (Conklin, B.R. et al. (1993) Cell 73:631-641). For example, the DNA fragment corresponding to the third intracellular loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the cell lysate by glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

For *in vitro* binding assays, cell extracts containing G proteins are prepared by extraction with 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM MgCl<sub>2</sub>, 20 mM CHAPS, 20% glycerol, 10 µg of both aprotinin and leupeptin, and 20 µl of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the supernatant is collected. 750 µg of cell extract is incubated with glutathione S-transferase (GST)  
30 fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound G subunits are detected by [<sup>32</sup>P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromphenol blue). The [<sup>32</sup>P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The  
35 separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat

dried milk, 50 mM Tris-HCl (pH 8.0), 2 mM CaCl<sub>2</sub>, 80 mM NaCl, 0.02% NaN<sub>3</sub>, and 0.2% Nonidet P-40) for 1 hour at room temperature, followed by incubation for 1.5 hours with G $\alpha$  subtype selective antibodies (1:500; Calbiochem-Novabiochem). After three washes, blots are incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin (1:2000, Cappel, Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

#### XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405).

- 10 Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

- 15 In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [<sup>3</sup>H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells.
- 20 Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of GCREC producing a 50% response level, where 100%
- 25 represents maximal incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

- In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or
- 35 without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M

perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

- 5 To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing  $1 \times 10^5$  cells/well and incubated with inositol-free media and [ $^3\text{H}$ ]myoinositol, 2  $\mu\text{Ci}$ /well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the
- 10 total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

#### XVIII. Identification of GCREC Ligands

- GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK
- 15 (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or  $\text{Ca}^{2+}$ . These may be measured directly using standard methods well known in the art,
- 20 or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate
- 25 Assay (NEN Life Sciences Products), or fluorescent  $\text{Ca}^{2+}$  indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins  $G_{\alpha 15/16}$  which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel
- 30 the signal transduction of the GCREC through a pathway involving phospholipase C and  $\text{Ca}^{2+}$  mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and  $G_{\alpha}$  protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways
- 35 are also modified so that the normal yeast response to the signal is converted to positive growth on



selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

5

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific  
10 embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1714538	1	1714538CD1	18	1714538CB1
3406743	2	3406743CD1	19	3406743CB1
3485895	3	3485895CD1	20	3485895CB1
7476102	4	7476102CD1	21	7476102CB1
2432942	5	2432942CD1	22	2432942CB1
4630911	6	4630911CD1	23	4630911CB1
7472432	7	7472432CD1	24	7472432CB1
7474977	8	7474977CD1	25	7474977CB1
7474848	9	7474848CD1	26	7474848CB1
7655614	10	7655614CD1	27	7655614CB1
6792419	11	6792419CD1	28	6792419CB1
7474790	12	7474790CD1	29	7474790CB1
7474816	13	7474816CD1	30	7474816CB1
7476172	14	7476172CD1	31	7476172CB1
7472141	15	7472141CD1	32	7472141CB1
7472137	16	7472137CD1	33	7472137CB1
7477934	17	7477934CD1	34	7477934CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	1714538CD1	g2865470	1.50E-217	[Homo sapiens] orphan G protein-coupled receptor; GPC-R (Tan, C. et al. (1998) Genomics 52:223-229)
2	3406743CD1	g10946201	0	[Homo sapiens] neuromedin U receptor 1
3	3485895CD1	g1902966	3.40E-16	[Mus sp.] oxytocin receptor (Kubota, Y. et al. (1996) Mol. Cell Endocrinol. 124:25-32)
4	7476102CD1	g13517983	0	[Homo sapiens] G-protein coupled receptor 91
5	2432942CD1	g767873	1.10E-47	[Rattus norvegicus] P2Y purinoceptor (Tokuyama, Y. et al. (1995) Biochem. Biophys. Res. Commun. 211:211-218)
6	4630911CD1	g202543	1.50E-12	[Rattus norvegicus] serotonin receptor
7	7472432CD1	g6006811	6.80E-73	[Mus musculus] serpentine receptor
8	7474977CD1	g13517962	0	[Homo sapiens] putative purinergic receptor
9	7474848CD1	g2231669	5.00E-45	[Homo sapiens] purinergic receptor P2Y9
10	7655614CD1	g2736345	4.00E-27	[Caenorhabditis elegans] contains similarity to G-coupled protein
11	6792419CD1	g310075	2.00E-151	[Rattus norvegicus] serotonin receptor
12	7474790CD1	g5019562	1.10E-37	[Homo sapiens] MAS-related G protein-coupled receptor
		g205314	9.70E-44	[Rattus norvegicus] mas oncogene (GPCR) encoded protein (Young, D. et al. (1988) Proc. Natl. Acad. Sci. USA 85:5339-5342)
		g189270	2.90E-147	[Homo sapiens] leukocyte platelet-activating factor receptor (GPCR) (Kunz, D. et al. (1992) J. Biol. Chem. 267:9101-9106)
		g4580924	2.30E-289	[Homo sapiens] endothelin receptor B delta 3 (Tsutsumi, M. et al. (1999) Gene 228:43-49)
		g2198745	2.10E-30	[Oryctolagus cuniculus] alpha 1a-adrenoceptor

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
13	7474816CD1	g11967419 g1055254	4.00E-41 2.70E-25	[Mus musculus] vomeronasal receptor VLRC3 [Rattus norvegicus] pheromone receptor VN6 (Dulac, C. and R. Axel (1995) Cell 83:195-206)
14	7476172CD1	g5809686 g3130153	2.60E-163 2.80E-117	[Carassius auratus] odorant receptor 5.24 [Takifugu rubripes] calcium2+ sensing receptor (Naito, T. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5178-5181)
15	7472141CD1	g7340541	1.90E-148	[Mus musculus] bM332P19.1 (novel 7 transmembrane receptor protein, rhodopsin family, olfactory receptor like) (mm17M1-12)
16	7472137CD1	g1256393	2.00E-59	[Rattus norvegicus] taste bud receptor protein TB 641 (Thomas, M.B. et al. (1996) Gene 178:1-5)
17	7477934CD1	g11692583 g12007423 g1016362	1.00E-116 2.00E-77 3.60E-69	[Mus musculus] odorant receptor M34 [Mus musculus] T2 olfactory receptor [Rattus norvegicus] OL1 receptor (Drutel, G. (1995) Receptors Channels 3:33-40)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1714538CD1	426	T366 T397 T282 S360	N7 N27 N41 N151 N196	Rhodopsin-like GPCR superfamily PR00237: P62-L86, T95-L116, F141-V163, H177-S198, V232-G255, V295-M319, H338-R364 ORPHAN G PROTEIN-COUPLED RECEPTOR PD027492: M24-K89 ORPHAN G PROTEIN-COUPLED RECEPTOR PD061371: S357-S426 RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE FAMILY PD000009: R87-P195 G-PROTEIN COUPLED RECEPTORS DM00013: I63-L371, Y67-R259, Y67-L258, L59-R272 G Protein Receptor: A147-V163 Transmembrane domain: M61-C82, A236-V254 7 transmembrane receptor (rhodopsin family) 7tm_1: G77-Y356 G-protein coupled receptor BL00237: F127-P166, F240-C251, R290-D316, G348-R364 G-protein coupled receptors signature: L139-V185	BLIMPS-PRINTS  BLAST-PRODOM  BLAST-PRODOM  BLAST-PRODOM  BLAST-DOMO  MOTIFS HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN
2	3406743CD1	259	S41 S131 S156 S242	N15 N27 N60	G-PROTEIN COUPLED RECEPTORS DM00013 P30559 32-344:R43-W176 Signal cleavage: M1-A58 Transmembrane domain: R45-C65 Rhodopsin-like GPCR super family PR00237: V44-C68, M82-S103, Q129-P151, A182-L205	BLAST-DOMO SPSCAN HMMER BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	3485895CD1	379	S111 T216 S222 S375 T55 S96 S142 S307	N53 N101 N104 N109 N213	G-protein coupled receptor	BLIMPS-BLOCKS
					BL00237D: N332-R348	BLIMPS-BLOCKS
					G-protein coupled receptors signature : Y145-I191	PROFILES SCAN
					Rhodopsin-like GPCR super family	BLIMPS-PRINTS
					PR00237: Y70-I94, S103-L124, L147-I169, 183-P204, S233-A256, P275-V299, Y322-R348	BLIMPS-PRINTS
					RECEPTOR COUPLED G-PROTEIN	BLAST-PRODROM
					TRANSMEMBRANE GLYCOPROTEIN	BLAST-PRODROM
					PHOSPHORYLATION LIPOPROTEIN	BLAST-PRODROM
					PALMITATE PD000009: S103-I206	MOTIFS
					G Protein Receptor: T153-I169	BLAST-DOMO
4	7476102CD1	396	S337 Y86	N34	G-PROTEIN COUPLED RECEPTORS	BLAST-DOMO
					DM00013: Y69-R356, K68-G344, L71-L351	BLAST-DOMO
					Transmembrane domain: I187-I206, M234-Y253, L276-M296, N319-L342	HMMER
					7 transmembrane receptor (rhodopsin family) 7tm_1: G85-Y340	HMMER-PFAM
					G-protein coupled receptor BL00237: W133-P172, F241-Y252, L270-M296	BLIMPS-BLOCKS
					Leucine Zipper: L80-L101	MOTIFS
					Rhodopsin-like GPCR superfamily	BLIMPS-PRINTS
					PR00237: V128-V150, A164-W185, T220-Y243	BLIMPS-PRINTS
					G-PROTEIN COUPLED RECEPTORS	BLAST-DOMO
					DM00013: S46-G333, L54-E196	BLAST-DOMO
5	7476102CD1	396	S337 Y86	N34	Signal cleavage: M1-A65	SPSCAN
					Transmembrane domain: L87-I106, P211-C231	HMMER
					Rhodopsin-like GPCR super family	BLIMPS-PRINTS
					PR00237B: H85-I106	BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	2432942CD1	528	S277 S29 S375 S449 S45 S505 S510 T23 T437 T93	N146 N147 N173 N179 N18 N394 N400 N58 N65	G-protein coupled receptor: PD000752: E243-K515 G-protein coupled receptor BL00649: A: G395-L422; B: G255-L300; C: C314-L339; D: G362-I386; G: S493-I518 Secretin-like GPCR superfamily PR00249: A: Y250-R274; B: R282-M306; C: A316-L339; D: V355-V380; F: D455-F475; G: L485-Q506 signal peptide: M1-T21 transmembrane domain: I251-R274, W470-F497 7 transmembrane receptor (Secretin family) 7tm_2: L245-A512 Latrophilin/CL-1-like GPS domain: GPS: Y185-A238, PUTATIVE PURINERGIC RECEPTOR P2Y10: PD055638: S16-V83 G-protein coupled receptor BL00237: A: W112-P151; B: F222-T233; C: G251-S277; D: N312-R328 Rhodopsin-like GPCR superfamily PR00237: A: I49-Y73; B: A82-R103; C: K126-L148; D: Y161-L182; E: M214-V237; F: A256-L280; G: H302-R328 transmembrane domain: A51-Y73, A82- Y106, I134-F152, I165-L183, V211- T233 7 transmembrane receptor (rhodopsin family): 7tm_1: G64-Y320	BLAST- PRODOR BLIMPS- BLOCKS BLIMPS- PRINTS HMMER HMMER HMMER-PFAM HMMER-PFAM BLAST- PRODOR BLIMPS- BLOCKS BLIMPS- PRINTS HMMER HMMER-PFAM
6	4630911CD1	361	S141 S239 S313 S323 S332 T185 T215 T233 T41 T79	N192 N350 N39		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7472432CD1	469	S207 S251 S264 S458 S217 S27 S289 T116 T163	N168 N357 N94	transmembrane domain: M74-W93, L221-L241, L360-M378 G-PROTEIN COUPLED RECEPTORS: DM00013 P31387 46-367: P46-T369 5HYDROXYTRYPTAMINE 5B RECEPTOR: PD027821: M1-H84 G-protein coupled receptors: BL00237A: R120-H159 BL00237B: F222-Y233 BL00237C: Q293-T319 BL00237D: N345-N361 Rhodopsin-like GPCR superfamily: PR00237A: L54-P78 PR00237B: P87-P108 PR00237C: H134-I156 PR00237D: A170-L191 PR00237E: A214-Y237 PR00237F: A298-I322 PR00237G: K335-N361 5-hydroxytryptamine 5B receptor: PR00519A: A3-P19 PR00519B: E20-P36 PR00519C: P36-V50 PR00519D: F196-R204 PR00519E: R246-V260 PR00519F: V262-V272 transmembrane domain: F48-L68, M300-L318 7 transmembrane receptor (rhodopsin family): 7tm_1: W69-Y353 signal_cleavage: M1-A65	HMMER BLAST-DOMO BLAST-PRODOR BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS HMMER HMMER-PFAM SPSCAN



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7474848CD1	330	T125, S284, S161, S173	N271	G-protein coupled receptors signature: F105-L153 Rhodopsin-like GPCR superfamily: PR00237A: I32-G56; PR00237B: F64-N85; PR00237C: M109-V131; PR00237D: L145-K166; PR00237E: L221-L245; PR00237G: H261-K287 G-PROTEIN COUPLED RECEPTOR, TRANSMEMBRANE GLYCOPROTEIN: PD013244; F193-E309 G-Protein_Receptor (0210.pdoc): A115-V131 G-PROTEIN COUPLED RECEPTORS: DM00013  P04201 26-296: I32-Q288 Transmembrane domain: I32-M59; C80-I97; Y113-W133; L145-I162; L191-C211; L218-G236 7 transmembrane receptor (rhodopsin family, 7tm_1): G47-Y279 G-protein coupled receptor: BL00237A: C95-P134; BL00237B: T30-A41; BL00237C: L216-Q242; BL00237D: N271-K287 Rhodopsin-like GPCR superfamily: PR00237A: A105-V129; PR00237B: T138-A159; PR00237C: S191-I213; PR00237D: R225-L246; PR00237E: S278-C301; PR00237F: T329-L353; PR00237G: S369-S395 G-protein coupled receptor: BL00237A: A177-P216; BL00237B: Y286-H297; BL00237C: F324-L350; BL00237D: N379-S395	PROFILES-SCAN BLIMPS-PRINTS BLAST-PRODROM MOTIFS BLAST-DOMO HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-BLOCKS
10	7655614CD1	494	S146, T266, T303, Y406	N54, N393		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical (Methods and Databases)
10					RHODOPSIN GTP-BINDING: PD062404: A249-A330 G-PROTEIN COUPLED RECEPTORS: DM00013 P47211 27-319: L108-N393 Transmembrane domain: L110-V129 7 transmembrane receptor (rhodopsin family; 7tm_1): G120-Y387 G protein receptor (O210.pdoc): V197-I213 G-protein coupled receptors signature: F189-G239	BLAST- PRODOM BLAST-DOMO HMMER HMMER-PFAM MOTIFS PROFILESCAN
11	6792419CD1	532	S14 S25 S36 S395 S43 S480 S497 S503 S526 T189 T308 T361 Y520	N149 N2 N209	transmembrane domain: T189-I206, V313-M335, W365-M390 7 transmembrane receptor (rhodopsin family): G208-L476 G_PROTEIN_RECEPTOR: BL00237A: W257-W296; BL00237B: F372-Y383; BL00237C: K406-S432; BL00237D: N468-K484 Endothelin-B receptor signature: PR00571A: Q92-G110; PR00571B: G110-T127; PR00571C: P128-W144; PR00571D: R154-T174; PR00571E: D256-V267; PR00571F: H348-A362; PR00571G: F498-N516; PR00571H: D517-S531 Endothelin receptor signature: PR00366A: C180-C199; PR00366B: R214-P226; PR00366C: S295-E311; PR00366D: P325-G339; PR00366E: F355-C373; PR00366F: T379-G396; PR00366G: M397-V411	HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					<p>Bombesin receptor signature: PR00358A: D244-F259; PR00358B: K265-S281; PR00358C: V479-C492</p> <p>G-protein coupled receptors signature: P268-I315</p> <p>G-Protein Receptor: I277-V293</p> <p>RECEPTOR G PROTEIN COUPLED TRANSMEMBRANE ETB GLYCOPROTEIN ENDOTHELIN B PRECURSOR PD010509: M91-R166</p> <p>RECEPTOR G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN SIGNAL ENDOTHELIN TYPE ETB PD004497: T167-C221</p> <p>RECEPTOR G PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN SIGNAL ENDOTHELIN TYPE B PD010472: L478-S532</p> <p>G-PROTEIN COUPLED RECEPTORS DM00013 P28088 94-400: E185-L491 DM00013 P21450 74-385: E185-C492 DM00013 A54126 63-374: E185-C492 DM00013 P32940 82-398: E185-L491</p> <p>signal peptide: M1-A67</p> <p>signal peptide: M1-A30</p> <p>transmembrane domain: T13-L33, A83- S103, S168-Y189, I378-W401</p> <p>7 transmembrane receptor (rhodopsin family): G25-Y431</p> <p>G-PROTEIN RECEPTOR: BL00237A: W74-P113; BL00237B: I178- Y189; BL00237C: C369-L395; BL00237D: Q423-K439</p>	<p>BLIMPS- PRINTS</p> <p>PROFILES- SCAN</p> <p>MOTIFS</p> <p>BLAST- PRODROM</p> <p>BLAST- PRODROM</p> <p>BLAST- PRODROM</p> <p>BLAST-DOMO</p> <p>SPSCAN</p> <p>HMMER</p> <p>HMMER</p> <p>HMMER-PFAM</p> <p>BLIMPS- BLOCKS</p>
12	7474790CD1	485	S210 S260 S273 S301 S308 S340 S356 S360 S460 T121 T302 T43 T437	N338		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical (Methods and Databases)
12					Rhodopsin-like GPCR superfamily signature: PR00237A: I10-Q34; PR00237E: T43-W64; PR00237C: T88-I110; PR00237D: R124-Y145; PR00237F: T170-F193; PR00237G: A374-L398; PR00237H: I413-K439 G-protein coupled receptors signature: L84-L129 G_Protein_Receptor: A94-I110 RECEPTOR COUPLED G PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PROTEIN FAMILY PD000009: Q34-W147	BLIMPS-PRINTS
13	7474816CD1	255		N145	G-PROTEIN COUPLED RECEPTORS DM000013: P18130 20-341: P2-Q199, K373-L446 P25100 90-417: A6-R208, K373-L446 P35368 39-364: I3-R198, N338-K440 P50407 76-400: A6-A201, I337-L446 transmembrane domain: Y37-L57 PHEROMONE RECEPTOR PD009900: L26-L247	HMMER BLAST- PRODOM
14	7476172CD1	881	S124 S130 S2 S303 S333 S636 S841 S879 T168 T169 T214 T407 T425 T461 T55 T613 T66	N212 N257 N285 N331 N38 N39 N405 N522 N545 N688 N74 N832 N865	transmembrane domains: I550-I569, F625-A642, Y657-W677, L705-F724 7 transmembrane receptor (metabotropic glutamate family): L548-N799 Receptor family ligand binding region: F15-I456	HMMER HMMER-PFAM HMMER-PFAM

Table 3 (cont.)

[illegible]

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15					<p>Olfactory receptor signature  PR00245A: M58-K79  PR00245B: F176-N190  PR00245C: L238-A253  PR00245D: I274-L285  PR00245E: T291-V305</p> <p>RECEPTOR OLFACATORY PROTEIN G-  PROTEIN COUPLED TRANSMEMBRANE  GLYCOPROTEIN MULTIGENE FAMILY  PD000921: L165-M246</p> <p>OLFACATORY RECEPTOR PROTEIN G-  PROTEIN COUPLED TRANSMEMBRANE  GLYCOPROTEIN MULTIGENE FAMILY  PD149621: V247-R303</p> <p>G-PROTEIN COUPLED RECEPTORS  DM00013 P23274 18-306: I25-R303  DM00013 P23275 17-306: V16-R303  DM00013 P30955 18-305: I25-V305  DM00013 P23266 17-306: I25-V305</p>	BLIMPS-PRINTS
16	7472137CD1	224	S135 S191 S50 S65 S76	N3	<p>Transmembrane domain:  V27-A46, I133-T159, M195-I212</p> <p>7 transmembrane receptor (rhodopsin family): G39-M144</p> <p>G-protein coupled receptors  signature: F100-F149</p> <p>G-protein coupled receptor domain  BL00237: R88-P127</p> <p>Olfactory receptor signature  PR00245: M57-K78, Y175-D189</p> <p>Melanocortin receptor family  PR00534: S49-L61, I124-S135  (P&lt;0.004)</p>	<p>BLAST-  PRODOM</p> <p>BLAST-  PRODOM</p> <p>BLAST-DOMO</p> <p>HMMER</p> <p>HMMER-PFAM</p> <p>PROFILESKAN</p> <p>BLIMPS-  BLOCKS</p> <p>BLIMPS-  PRINTS</p> <p>BLIMPS-  PRINTS</p>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16					OLFACTORY RECEPTOR PROTEIN PD000921: L164-L224	BLAST- PRODOM
17	7477934CD1	326	S186 S20 S235 S289 S36 S65 T190 T191	N153 N5	G-PROTEIN COUPLED RECEPTORS DM00013 P30955 18-305: I24-L224 7 transmembrane receptor (rhodopsin family) 7tm_1: E39-Y288 G-PROTEIN COUPLED RECEPTORS DM00013 A57069 15-304: F33-L303 RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: F166-I244 OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: V245-F307 G-protein coupled receptor BL00237: R88-P127, L205-Y216, T280-L296 Olfactory receptor signature PR00245: F236-G251, L272-L283, S289-L303, M57-K78, F175-D189 G-protein coupled receptors signature: F100-G145 transmembrane domain: L25-L46 G-Protein_Receptor: G108-I124	BLAST-DOMO BLAST-DOMO HMMER-PFAM BLAST-DOMO BLAST- PRODOM BLIMPS- BLOCKS BLIMPS- PRINTS PROFILESCAN HMMER MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
18	1714538CB1	2374	1-1852, 2139-2374, 1-234, 818-1572, 532-1151, 68-691	71687857V1	2141	2287
				71688177V1	986	1547
				71688995V1	1246	1843
				71687846V1	1684	2287
				1714538H1 (UCMCNOT02)	4	289
19	3406743CB1	813	1-168, 321-725	GNN:g6855250_000025_002_edit	87	1365
				4516658H1 (SINJNOT03)	2113	2374
				FL210313_00001	1	813
20	3485895CB1	1542	1-334, 626-883, 1518-1542	71296533V1	926	1542
				71755255V1	525	1186
				60149346D1	87	518
				GNN.g6479070_edit	1	300
				71295925V1	492	1180
21	7476102CB1	1191	1168-1191, 1-1097	GBI.g7684460_edit	1	1191
22	2432942CB1	3360	1-1940, 3316-3360, 2803-2843	2432516H1 (BRAVUNT02)	1	238
				70572441V1	1447	2041
				70254651V1	997	1498
				8116752H1 (TONSDIC01)	682	1314
				71733263V1	300	966
				71873820V1	2675	3360
				55046039J2	131	943
				6544942F6 (LNCN02)	2018	2827
				71730213V1	2145	2854
				70815928V1	1502	2139
23	4630911CB1	1660	1-1260	2204242T6 (SPINFET02)	1263	1660
				71047664V1	924	1516
				7706542J1 (UTRETUE01)	245	563
				GBI.g8112480_edit	1	1125
				6589031F8 (TLYMUNT03)	459	1158



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
24	7472432CB1	2603	1-24, 2540-2603, 838-2005	6975661H1 (BRAHTDR04) 55037225J2 7341002H1 (COLNDIN02) 7059069H1 (BRALNON02) FL7472432CB1_00001 2964391F6 (SCORNOT04) g7137674_edit GBI.g8081257_edit	1657 1 2336 1819 239 1353 1 1	2249 393 2603 2477 1648 1728 1119 1018
25	7474977CB1	1119	1-1119			
26	7474848CB1	1018	1-140, 262-344, 546-571, 992-1018			
35	7655614CB1.comp	2177	1537-2177, 1-39, 857-1072, 372-472	7655614H1 (UTREDME06) FL7655614_g7406476_g156725 7655614J1 (UTREDME06)	653 705 1	1240 2177 710
28	6792419CB1	1632	1-234, 1-695, 752-809	6314367T6 (NERDPTDN03) 6110713F8 (MCLDPTXT03) g4580923_CD 70619934V1 GNN.g6759183_000017_002_edit	811 161 1 1048 1	1620 968 1632 1624 1458
29	7474790CB1	1458	1-103, 370-1458			
30	7474816CB1	1015	1-489, 564-1015	55078072J1 FL7474816_g8080066_000028_g1055254 FL7476172_g7684554_000005_g5809686	1 248 1	596 1015 2781
31	7476172CB1	2781	888-2781, 1-252, 1-393, 687-2781			
32	7472141CB1	1267	1-449, 1-938	g7940222_edit	1	1267
33	7472137CB1	1559	1268-1559, 1-1043	FL7472137_g8118970_000006_g1256393 55033677H1 (FLP4X0017) GBI.g8568403_000012_edit	624 1 1	1559 877 981
34	7477934CB1	981	43-92, 909-981			

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
18	1714538CB1	BRSTNON02
19	3406743CB1	PROSTUS08
20	3485895CB1	KIDNNOT20
22	2432942CB1	BRAVUNT02
23	4630911CB1	SPLANFET02
24	7472432CB1	SCORNOT04
27	7655614CB1	PITUDIR01
28	6792419CB1	NERDTN03

Table 6

Library	Vector	Library Description
BRAVUNT02	PSPORT1	Library was constructed using polyA RNA isolated from separate populations of unstimulated astrocytes. The RNA was pooled for polyA RNA isolation and library construction.
BRSTN02	pINCY	This normalized breast tissue library was constructed from 6.2 million independent clones from a pool of two libraries from two different donors. Starting RNA was made from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammaplasty (donor A), and from breast tissue removed from a 60-year-old Caucasian female during a bilateral reduction mammaplasty (donor B). Pathology indicated normal breast parenchyma, bilaterally (A) and bilateral mammary hypertrophy (B). Patient history included hypertrophy of breast, obesity, lumbago, and glaucoma (A) and joint pain in the shoulder, thyroid cyst, colon cancer, normal delivery and cervical cancer (B). Family history included cataract, osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, and alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes (A) and cerebrovascular accident, atherosclerotic coronary artery disease, colon cancer, type II diabetes, hyperlipidemia, depressive disorder, and Alzheimer's Disease. The library was normalized in two rounds using conditions adapted from Soares et al. (1994) PNAS 91:9228-9232 and Bonaldo et al. (1996) Genome Research 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
KIDNNOT20	pINCY	Library was constructed using RNA isolated from left kidney tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma. Family history included atherosclerotic coronary artery disease.
PITUDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from pituitary gland tissue removed from a 70-year-old female who died from metastatic adenocarcinoma. Pathology for the brain indicated moderate Alzheimer disease and mild carotid and cerebral atherosclerosis. The cerebral hemispheres, frontal and temporal lobes, white matter, and hippocampus showed mild atrophy, bilaterally. There were numerous neurofibrillary tangles, neuritic and diffuse amyloid plaques deposited throughout most neocortical areas.
SPLNFET02	pINCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
NERDTN03	pINCY	<p>This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone), hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytobom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al. (1994) PNAS 91:9228-9232 and Bonaldo et al. (1996) Genome Research 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.</p>
PROSTUS08	PT7T3	<p>This subtracted library was constructed using 2.36 million clones from a prostate tumor library and was subjected to one round of subtractive hybridization with 448,000 clones from a normal prostate library. The starting library for subtraction was constructed using RNA isolated from a prostate tumor removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) Adenofibrous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991) NAR 19:1954 and Bonaldo et al. (1996) Genome Research 6:791.</p>
SCORNOT04	pINCY	<p>Library was constructed using RNA isolated from cervical spinal cord tissue removed from a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.</p>

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater, Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

Tissues	Polynucleotide SEQ ID NO:
Breast, Fat, Skin	33
Muscle, Bone, Synovium, Connective tissue	-
Pancreas, Liver, Gallbladder	+
Brain: Amygdala, Thalamus, Hippocampus, Entorhinal cortex, Archacocortex	-
Brain: Striatum, Caudate nucleus, Putamen, Dentate nucleus, Globus pallidus, Substantia innominata, Ralphhe magnus	-
Kidney, Fetal colon, Small intestine, Ileum, Esophagus	-
Fetal heart, Aorta, Coronary artery	-
Fetal lung, Adult lung	-
Placenta, Prostate, Uterus	-
Olfactory bulb	-

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-17,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and
  - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-  
17.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:18-34.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and



b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

5 11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34,

10 c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a  
15 polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides  
20 comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

25

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction  
30 amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected  
5 from the group consisting of SEQ ID NO:1-17.

18. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 16.

10

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

15

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of  
20 functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- 25 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

30

24. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim  
35 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

5

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- 10 b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a
- 15 compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 20 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

25

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific
- 30 hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the
- 35 amount of hybridization complex in an untreated biological sample, wherein a difference in the

amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of GCREC in a  
5 biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

10

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- 15 d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

15

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

20

32. A method of diagnosing a condition or disease associated with the expression of GCREC in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

25

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of GCREC in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

30

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

- b) isolating antibodies from said animal; and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

5

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

10 38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- 15 b) isolating antibody producing cells from the animal;
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;
- d) culturing the hybridoma cells; and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide
- 20 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant

30 immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 in a sample, comprising the steps of:

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific

binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 in the sample.

5

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

10 b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15 46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25 51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:18.
63. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:19.
64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.
65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.
66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.
67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.
68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.
69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:26.

5

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
10 NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:29.

15 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:31.

20

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
25 NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:34.



<110> INCYTE GENOMICS, INC.  
 LAL, Preeti  
 BAUGHN, Mariah R.  
 HAFALIA, April J. A.  
 NGUYEN, Dannel B.  
 GANDHI, Ameena R.  
 KALLICK, Deborah A.  
 GRIFFIN, Jennifer A.  
 YUE, Henry  
 KHAN, Farrah A.  
 PATTERSON, Chandra  
 LU, Dyung Aina M.  
 TRIBOULEY, Catherine M.  
 LU, Yan  
 WALIA, Narinder K.  
 GRAUL, Richard  
 YAO, Monique G.  
 YANG, Junming  
 RAMKUMAR, Jayalaxmi  
 AU-YOUNG, Janice  
 ELLIOTT, Vicki S.  
 HERNANDEZ, Roberto  
 WALSH, Roderick T.  
 BOROWSKY, Mark L.  
 THORNTON, Michael  
 HE, Ann

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0131 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/212,483; 60/213,954; 60/215,209; 60/216,595; 60/218,936;  
 60/219,154; 60/220,141

<151> 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14;  
 2000-07-19; 2000-07-21

<160> 35

<170> PERL Program

<210> 1

<211> 426

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1714538CD1

<400> 1

Met	Thr	Pro	Leu	Cys	Leu	Asn	Cys	Ser	Val	Leu	Pro	Gly	Asp	Leu
1				5					10					15
Tyr	Pro	Gly	Gly	Ala	Arg	Asn	Pro	Met	Ala	Cys	Asn	Gly	Ser	Ala
				20					25					30
Ala	Arg	Gly	His	Phe	Asp	Pro	Glu	Asp	Leu	Asn	Leu	Thr	Asp	Glu
				35					40					45
Ala	Leu	Arg	Leu	Lys	Tyr	Leu	Gly	Pro	Gln	Gln	Thr	Glu	Leu	Phe
				50					55					60
Met	Pro	Ile	Cys	Ala	Thr	Tyr	Leu	Leu	Ile	Phe	Val	Val	Gly	Ala
				65					70					75

Val	Gly	Asn	Gly	Leu	Thr	Cys	Leu	Val	Ile	Leu	Arg	His	Lys	Ala	
				80					85					90	
Met	Arg	Thr	Pro	Thr	Asn	Tyr	Tyr	Leu	Phe	Ser	Leu	Ala	Val	Ser	
				95					100					105	
Asp	Leu	Leu	Val	Leu	Leu	Val	Gly	Leu	Pro	Leu	Glu	Leu	Tyr	Glu	
				110					115					120	
Met	Trp	His	Asn	Tyr	Pro	Phe	Leu	Leu	Gly	Val	Gly	Gly	Cys	Tyr	
				125					130					135	
Phe	Arg	Thr	Leu	Leu	Phe	Glu	Met	Val	Cys	Leu	Ala	Ser	Val	Leu	
				140					145					150	
Asn	Val	Thr	Ala	Leu	Ser	Val	Glu	Arg	Tyr	Val	Ala	Val	Val	His	
				155					160					165	
Pro	Leu	Gln	Ala	Arg	Ser	Met	Val	Thr	Arg	Ala	His	Val	Arg	Arg	
				170					175					180	
Val	Leu	Gly	Ala	Val	Trp	Gly	Leu	Ala	Met	Leu	Cys	Ser	Leu	Pro	
				185					190					195	
Asn	Thr	Ser	Leu	His	Gly	Ile	Arg	Gln	Leu	His	Val	Pro	Cys	Arg	
				200					205					210	
Gly	Pro	Val	Pro	Asp	Ser	Ala	Val	Cys	Met	Leu	Val	Arg	Pro	Arg	
				215					220					225	
Ala	Leu	Tyr	Asn	Met	Val	Val	Gln	Thr	Thr	Ala	Leu	Leu	Phe	Phe	
				230					235					240	
Cys	Leu	Pro	Met	Ala	Ile	Met	Ser	Val	Leu	Cys	Leu	Leu	Val	Gly	
				245					250					255	
Leu	Arg	Leu	Arg	Arg	Glu	Arg	Leu	Leu	Leu	Met	Gln	Glu	Ala	Lys	
				260					265					270	
Gly	Arg	Gly	Ser	Ala	Ala	Ala	Arg	Ser	Arg	Tyr	Thr	Cys	Arg	Leu	
				275					280					285	
Gln	Gln	His	Asp	Arg	Gly	Arg	Gly	Gln	Val	Thr	Lys	Met	Leu	Phe	
				290					295					300	
Val	Leu	Val	Val	Val	Phe	Gly	Ile	Cys	Trp	Ala	Pro	Phe	His	Ala	
				305					310					315	
Asp	Arg	Val	Met	Trp	Ser	Val	Val	Ser	Gln	Trp	Thr	Asp	Gly	Leu	
				320					325					330	
His	Leu	Ala	Phe	Gln	His	Val	His	Val	Ile	Ser	Gly	Ile	Phe	Phe	
				335					340					345	
Tyr	Leu	Gly	Ser	Ala	Ala	Asn	Pro	Val	Leu	Tyr	Ser	Leu	Met	Ser	
				350					355					360	
Ser	Arg	Phe	Arg	Glu	Thr	Phe	Gln	Glu	Ala	Leu	Cys	Leu	Gly	Ala	
				365					370					375	
Cys	Cys	His	Arg	Leu	Arg	Pro	Arg	His	Ser	Ser	His	Ser	Leu	Ser	
				380					385					390	
Arg	Met	Thr	Thr	Gly	Ser	Thr	Leu	Cys	Asp	Val	Gly	Ser	Leu	Gly	
				395					400					405	
Ser	Trp	Val	His	Pro	Leu	Ala	Gly	Asn	Asp	Gly	Pro	Glu	Ala	Gln	
				410					415					420	
Gln	Glu	Thr	Asp	Pro	Ser										
				425											

&lt;210&gt; 2

&lt;211&gt; 259

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3406743CD1

&lt;400&gt; 2

Met	Glu	Asp	Leu	Phe	Ser	Pro	Ser	Ile	Leu	Pro	Pro	Ala	Pro	Asn	
1				5					10					15	
Ile	Ser	Val	Pro	Ile	Leu	Leu	Gly	Trp	Gly	Leu	Asn	Leu	Thr	Leu	
				20					25					30	

[illegible]

<210> 3

<211> 379

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3485895CD1

<400> 3

Met	Gln	Ile	Gly	His	Leu	Ala	Gln	His	Trp	Val	Arg	Ser	Tyr	Ile
1				5					10					15
Gln	Leu	Leu	Ala	Glu	Phe	Leu	Ser	Arg	Asp	Gln	Val	Phe	Gln	Gln
				20					25					30
Asn	Gly	Tyr	Gly	Leu	Thr	Gln	Gln	Asn	Leu	Leu	Asn	Asn	Tyr	Asp
				35					40					45
Met	Leu	Gly	Ile	Met	Ala	Trp	Asn	Ala	Thr	Cys	Lys	Asn	Trp	Leu
				50					55					60
Ala	Ala	Glu	Ala	Ala	Leu	Glu	Lys	Tyr	Tyr	Leu	Ser	Ile	Phe	Tyr
				65					70					75
Gly	Ile	Glu	Phe	Val	Val	Gly	Val	Leu	Gly	Asn	Thr	Ile	Val	Val
				80					85					90
Tyr	Gly	Tyr	Ile	Phe	Ser	Leu	Lys	Asn	Trp	Asn	Ser	Ser	Asn	Ile
				95					100					105
Tyr	Leu	Phe	Asn	Leu	Ser	Val	Ser	Asp	Leu	Ala	Phe	Leu	Cys	Thr
				110					115					120
Leu	Pro	Met	Leu	Ile	Arg	Ser	Tyr	Ala	Asn	Gly	Asn	Trp	Ile	Tyr
				125					130					135
Gly	Asp	Val	Leu	Cys	Ile	Ser	Asn	Arg	Tyr	Val	Leu	His	Ala	Asn
				140					145					150

```

Leu Tyr Thr Ser Ile Leu Phe Leu Thr Phe Ile Ser Ile Asp Arg
155 160 165
Tyr Leu Ile Ile Lys Tyr Pro Phe Arg Glu His Leu Leu Gln Lys
170 175 180
Lys Glu Phe Ala Ile Leu Ile Ser Leu Ala Ile Trp Val Leu Val
185 190 195
Thr Leu Glu Leu Leu Pro Ile Leu Pro Leu Ile Asn Pro Val Ile
200 205 210
Thr Asp Asn Gly Thr Thr Cys Asn Asp Phe Ala Ser Ser Gly Asp
215 220 225
Pro Asn Tyr Asn Leu Ile Tyr Ser Met Cys Leu Thr Leu Leu Gly
230 235 240
Phe Leu Ile Pro Leu Phe Val Met Cys Phe Phe Tyr Tyr Lys Ile
245 250 255
Ala Leu Phe Leu Lys Gln Arg Asn Arg Gln Val Ala Thr Ala Leu
260 265 270
Pro Leu Glu Lys Pro Leu Asn Leu Val Ile Met Ala Val Val Ile
275 280 285
Phe Ser Val Leu Phe Thr Pro Tyr His Val Met Arg Asn Val Arg
290 295 300
Ile Ala Ser Arg Leu Gly Ser Trp Lys Gln Tyr Gln Cys Thr Gln
305 310 315
Val Val Ile Asn Ser Phe Tyr Ile Val Thr Arg Pro Leu Ala Phe
320 325 330
Leu Asn Ser Val Ile Asn Pro Val Phe Tyr Phe Leu Leu Gly Asp
335 340 345
His Phe Arg Asp Met Leu Met Asn Gln Leu Arg His Asn Phe Lys
350 355 360
Ser Leu Thr Ser Phe Ser Arg Trp Ala His Glu Leu Leu Leu Ser
365 370 375
Phe Arg Glu Lys

```

&lt;210&gt; 4

&lt;211&gt; 396

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476102CD1

&lt;400&gt; 4

```

Met Gly Asp Glu Leu Ala Pro Cys Pro Val Gly Thr Thr Ala Trp
1 5 10 15
Pro Ala Leu Ile Gln Leu Ile Ser Lys Thr Pro Cys Met Pro Gln
20 25 30
Ala Ala Ser Asn Thr Ser Leu Gly Leu Gly Asp Leu Arg Val Pro
35 40 45
Ser Ser Met Leu Tyr Trp Leu Phe Leu Pro Ser Ser Leu Leu Ala
50 55 60
Ala Ala Thr Leu Ala Val Ser Pro Leu Leu Leu Val Thr Ile Leu
65 70 75
Arg Asn Gln Arg Leu Arg Gln Glu Pro His Tyr Leu Leu Pro Ala
80 85 90
Asn Ile Leu Leu Ser Asp Leu Ala Tyr Ile Leu Leu His Met Leu
95 100 105
Ile Ser Ser Ser Ser Leu Gly Gly Trp Glu Leu Gly Arg Met Ala
110 115 120
Cys Gly Ile Leu Thr Asp Ala Val Phe Ala Ala Cys Thr Ser Thr
125 130 135
Ile Leu Ser Phe Thr Ala Ile Val Leu His Thr Tyr Leu Ala Val
140 145 150

```

```

Ile His Pro Leu Arg Tyr Leu Ser Phe Met Ser His Gly Ala Ala
155 160 165
Trp Lys Ala Val Ala Leu Ile Trp Leu Val Ala Cys Cys Phe Pro
170 175 180
Thr Phe Leu Ile Trp Leu Ser Lys Trp Gln Asp Ala Gln Leu Glu
185 190 195
Glu Gln Gly Ala Ser Tyr Ile Leu Pro Pro Ser Met Gly Thr Gln
200 205 210
Pro Gly Cys Gly Leu Leu Val Ile Val Thr Tyr Thr Ser Ile Leu
215 220 225
Cys Val Leu Phe Leu Cys Thr Ala Leu Ile Ala Asn Cys Phe Trp
230 235 240
Arg Ile Tyr Ala Glu Ala Lys Thr Ser Gly Ile Trp Gly Gln Gly
245 250 255
Tyr Ser Arg Ala Arg Gly Thr Leu Leu Ile His Ser Val Leu Ile
260 265 270
Thr Leu Tyr Val Ser Thr Gly Val Val Phe Ser Leu Asp Met Val
275 280 285
Leu Thr Arg Tyr His His Ile Asp Ser Gly Thr His Thr Trp Leu
290 295 300
Leu Ala Ala Asn Ser Glu Val Leu Met Met Leu Pro Arg Ala Met
305 310 315
Leu Thr Tyr Leu Tyr Leu Leu Arg Tyr Arg Gln Leu Leu Gly Met
320 325 330
Val Arg Gly His Leu Pro Ser Arg Arg His Gln Ala Ile Phe Thr
335 340 345
Ile Ser Cys Cys Trp Glu Pro Leu Phe Tyr Phe Ser Leu Ile Leu
350 355 360
Thr Thr Leu Gly Val Asp Ile Ile Pro Leu Cys Val Glu Thr Thr
365 370 375
Phe Cys Leu Ser Ile His Leu Ser Met Asp Arg Leu Gly Cys Thr
380 385 390
Thr Phe Gly Tyr Cys Glu
395

```

&lt;210&gt; 5

&lt;211&gt; 528

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2432942CD1

&lt;400&gt; 5

```

Met Asp His Cys Gly Ala Leu Phe Leu Cys Leu Cys Leu Leu Thr
1 5 10 15
Leu Gln Asn Ala Thr Thr Glu Thr Trp Glu Leu Leu Ser Tyr
20 25 30
Met Glu Asn Met Gln Val Ser Arg Gly Arg Ser Ser Val Phe Ser
35 40 45
Ser Arg Gln Leu His Gln Leu Glu Gln Met Leu Leu Asn Thr Ser
50 55 60
Phe Pro Gly Tyr Asn Leu Thr Leu Gln Thr Pro Thr Ile Gln Ser
65 70 75
Leu Ala Phe Lys Leu Ser Cys Asp Phe Ser Gly Leu Ser Leu Thr
80 85 90
Ser Ala Thr Leu Lys Arg Val Pro Gln Ala Gly Gly Gln His Ala
95 100 105
Arg Gly Gln His Ala Met Gln Phe Pro Ala Glu Leu Thr Arg Asp
110 115 120
Ala Cys Lys Thr Arg Pro Arg Glu Leu Arg Leu Ile Cys Ile Tyr
125 130 135

```

Phe	Ser	Asn	Thr	His	Phe	Phe	Lys	Asp	Glu	Asn	Asn	Ser	Ser	Leu
				140					145					150
Leu	Asn	Asn	Tyr	Val	Leu	Gly	Ala	Gln	Leu	Ser	His	Gly	His	Val
				155					160					165
Asn	Asn	Leu	Arg	Asp	Pro	Val	Asn	Ile	Ser	Phe	Trp	His	Asn	Gln
				170					175					180
Ser	Leu	Glu	Gly	Tyr	Thr	Leu	Thr	Cys	Val	Phe	Trp	Lys	Glu	Gly
				185					190					195
Ala	Arg	Lys	Gln	Pro	Trp	Gly	Gly	Trp	Ser	Pro	Glu	Gly	Cys	Arg
				200					205					210
Thr	Glu	Gln	Pro	Ser	His	Ser	Gln	Val	Leu	Cys	Arg	Cys	Asn	His
				215					220					225
Leu	Thr	Tyr	Phe	Ala	Val	Leu	Met	Gln	Leu	Ser	Pro	Ala	Leu	Val
				230					235					240
Pro	Ala	Glu	Leu	Leu	Ala	Pro	Leu	Thr	Tyr	Ile	Ser	Leu	Val	Gly
				245					250					255
Cys	Ser	Ile	Ser	Ile	Val	Ala	Ser	Leu	Ile	Thr	Val	Leu	Leu	His
				260					265					270
Phe	His	Phe	Arg	Lys	Gln	Ser	Asp	Ser	Leu	Thr	Arg	Ile	His	Met
				275					280					285
Asn	Leu	His	Ala	Ser	Val	Leu	Leu	Leu	Asn	Ile	Ala	Phe	Leu	Leu
				290					295					300
Ser	Pro	Ala	Phe	Ala	Met	Ser	Pro	Val	Pro	Gly	Ser	Ala	Cys	Thr
				305					310					315
Ala	Leu	Ala	Ala	Ala	Leu	His	Tyr	Ala	Leu	Leu	Ser	Cys	Leu	Thr
				320					325					330
Trp	Met	Ala	Ile	Glu	Gly	Phe	Asn	Leu	Tyr	Leu	Leu	Leu	Gly	Arg
				335					340					345
Val	Tyr	Asn	Ile	Tyr	Ile	Arg	Arg	Tyr	Val	Phe	Lys	Leu	Gly	Val
				350					355					360
Leu	Gly	Trp	Gly	Ala	Pro	Ala	Leu	Leu	Val	Leu	Leu	Ser	Leu	Ser
				365					370					375
Val	Lys	Ser	Ser	Val	Tyr	Gly	Pro	Cys	Thr	Ile	Pro	Val	Phe	Asp
				380					385					390
Ser	Trp	Glu	Asn	Gly	Thr	Gly	Phe	Gln	Asn	Met	Ser	Ile	Cys	Trp
				395					400					405
Val	Arg	Ser	Pro	Val	Val	His	Ser	Val	Leu	Val	Met	Gly	Tyr	Gly
				410					415					420
Gly	Leu	Thr	Ser	Leu	Phe	Asn	Leu	Val	Val	Leu	Ala	Trp	Ala	Leu
				425					430					435
Trp	Thr	Leu	Arg	Arg	Leu	Arg	Glu	Arg	Ala	Asp	Ala	Pro	Ser	Val
				440					445					450
Arg	Ala	Cys	His	Asp	Thr	Val	Thr	Val	Leu	Gly	Leu	Thr	Val	Leu
				455					460					465
Leu	Gly	Thr	Thr	Trp	Ala	Leu	Ala	Phe	Phe	Ser	Phe	Gly	Val	Phe
				470					475					480
Leu	Leu	Pro	Gln	Leu	Phe	Leu	Phe	Thr	Ile	Leu	Asn	Ser	Leu	Tyr
				485					490					495
Gly	Phe	Phe	Leu	Phe	Leu	Trp	Phe	Cys	Ser	Gln	Arg	Cys	Arg	Ser
				500					505					510
Glu	Ala	Glu	Ala	Lys	Ala	Gln	Ile	Glu	Ala	Phe	Ser	Ser	Ser	Gln
				515					520					525
Thr	Thr	Gln												

&lt;210&gt; 6

&lt;211&gt; 361

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4630911CD1

&lt;400&gt; 6 -

```

Met Asn Arg Lys Gln Leu Glu Val Ile Pro Met Ile Trp Val Ile
 1          5          10          15
Ser Leu Asp Lys Ser Lys Gly Thr Phe His Ile Gly Ser His Phe
          20          25          30
Thr Ser Thr Ser Leu Ile Phe Ser Asn Cys Thr Asn Thr Asp Phe
          35          40          45
Arg Tyr Phe Ile Tyr Ala Val Thr Tyr Thr Val Ile Leu Val Pro
          50          55          60
Gly Leu Ile Gly Asn Ile Leu Ala Leu Trp Val Phe Tyr Gly Tyr
          65          70          75
Met Lys Glu Thr Lys Arg Ala Val Ile Phe Met Ile Asn Leu Ala
          80          85          90
Ile Ala Asp Leu Leu Gln Val Leu Ser Leu Pro Leu Arg Ile Phe
          95          100          105
Tyr Tyr Leu Asn His Asp Trp Pro Phe Gly Pro Gly Leu Cys Met
          110          115          120
Phe Cys Phe Tyr Leu Lys Tyr Val Asn Met Tyr Ala Ser Ile Tyr
          125          130          135
Phe Leu Val Cys Ile Ser Val Arg Arg Phe Trp Phe Leu Met Tyr
          140          145          150
Pro Phe Arg Phe His Asp Cys Lys Gln Lys Tyr Asp Leu Tyr Ile
          155          160          165
Ser Ile Ala Gly Trp Leu Ile Ile Cys Leu Ala Cys Val Leu Phe
          170          175          180
Pro Leu Leu Arg Thr Ser Asp Asp Thr Pro Gly Asn Arg Thr Lys
          185          190          195
Cys Phe Val Asp Leu Pro Thr Arg Asn Val Asn Leu Ala Gln Ser
          200          205          210
Val Val Met Met Thr Ile Gly Glu Leu Ile Gly Phe Val Thr Pro
          215          220          225
Leu Leu Ile Val Leu Tyr Cys Thr Trp Lys Thr Val Leu Ser Leu
          230          235          240
Gln Asp Lys Tyr Pro Met Ala Gln Asp Leu Gly Glu Lys Gln Lys
          245          250          255
Ala Leu Lys Met Ile Leu Thr Cys Ala Gly Val Phe Leu Ile Cys
          260          265          270
Phe Ala Pro Tyr His Phe Ser Phe Pro Leu Asp Phe Leu Val Lys
          275          280          285
Ser Asn Glu Ile Lys Ser Cys Leu Ala Arg Arg Val Ile Leu Ile
          290          295          300
Phe His Ser Val Ala Leu Cys Leu Ala Ser Leu Asn Ser Cys Leu
          305          310          315
Asp Pro Val Ile Tyr Tyr Phe Ser Thr Asn Glu Phe Arg Arg Arg
          320          325          330
Leu Ser Arg Gln Asp Leu His Asp Ser Ile Gln Leu His Ala Lys
          335          340          345
Ser Phe Val Ser Asn His Thr Ala Ser Thr Met Thr Pro Glu Leu
          350          355          360
Cys

```

&lt;210&gt; 7

&lt;211&gt; 469

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472432CD1

&lt;400&gt; 7

```

Met Ala Phe Leu Met His Leu Leu Val Cys Val Phe Gly Met Gly

```

1	-	5		10		15
Ser	Trp	Val	Thr	Ile	Asn	Gly
20						
Val	Met	Glu	Leu	Pro	Glu	Gly
35						
Val	Val	Ile	Gln	Leu	Ala	Asn
50						
Leu	His	His	Phe	Arg	Pro	Ser
65						
Phe	Thr	Leu	Leu	Gly	Val	Gly
80						
Phe	Leu	Trp	Asn	Met	Thr	Ser
95						
Ile	Ala	Phe	Leu	Val	Leu	Thr
110						
Thr	Ser	Ser	Val	Thr	Phe	Leu
125						
Tyr	Tyr	Leu	Thr	Thr	Phe	Phe
140						
Leu	Pro	Ala	Leu	Val	Ala	Leu
155						
Cys	Val	Asn	Val	Thr	Glu	Ile
170						
Pro	Thr	Arg	Glu	Thr	Asp	Ile
185						
Val	Ser	Ala	Leu	Pro	Gly	Met
200						
Ser	Arg	Tyr	Leu	Pro	Ala	His
215						
Leu	Leu	Ser	Ile	Met	Met	Ala
230						
Leu	Gln	Arg	Gln	Pro	Arg	Cys
245						
Leu	Asn	Asp	Gln	Val	Thr	Leu
260						
Asn	Asp	Leu	Gly	Pro	Ala	Gly
275						
Gly	Tyr	Leu	Glu	Glu	Lys	Ala
290						
Ala	Phe	Ile	Tyr	Thr	Leu	Val
305						
Gly	Met	Leu	Pro	Ser	Val	Gln
320						
Pro	Val	Ala	Tyr	His	Leu	Ala
335						
Pro	Leu	Ala	Ser	Leu	Val	Ser
350						
Leu	Phe	Leu	Gly	Val	Leu	Ser
365						
Tyr	Asn	Met	Ala	Met	Ala	Val
380						
Gly	His	Trp	Gly	Gly	Glu	Val
395						
Phe	Ser	Gly	Cys	Leu	Ser	Tyr
410						
Leu	Arg	Asp	Leu	Ser	Arg	Ser
425						
Val	Gln	Leu	Gly	Ser	Leu	Leu
440						
Val	Asn	Val	Leu	Arg	Leu	Phe
455						
His	Cys	Pro	Ala			



<210> 8 -  
 <211> 372  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474977CD1

<400> 8  
 Met Glu Ala Ala Ser Leu Ser Val Ala Thr Ala Gly Val Ala Leu  
 1 5 10 15  
 Ala Leu Gly Pro Glu Thr Ser Ser Gly Thr Pro Ser Pro Arg Gly  
 20 25 30  
 Ile Leu Gly Ser Thr Pro Ser Gly Ala Val Leu Pro Gly Arg Gly  
 35 40 45  
 Pro Pro Phe Ser Val Phe Thr Val Leu Val Val Thr Leu Leu Val  
 50 55 60  
 Leu Leu Ile Ala Ala Thr Phe Leu Trp Asn Leu Leu Val Pro Val  
 65 70 75  
 Thr Ile Pro Arg Val Arg Ala Phe His Arg Val Pro His Asn Leu  
 80 85 90  
 Val Ala Ser Thr Ala Val Ser Asp Glu Leu Val Ala Ala Leu Ala  
 95 100 105  
 Met Pro Pro Ser Leu Ala Ser Glu Leu Ser Thr Gly Arg Arg Arg  
 110 115 120  
 Leu Leu Gly Arg Ser Leu Cys His Val Trp Ile Ser Phe His Val  
 125 130 135  
 Leu Cys Cys Pro Ala Gly Leu Gly Asn Val Ala Ala Ile Ala Leu  
 140 145 150  
 Gly Arg Asp Gly Ala Ile Thr Arg His Leu Gln His Thr Leu Arg  
 155 160 165  
 Thr Arg Ser Arg Ala Ser Leu Leu Met Ile Ala Leu Thr Arg Val  
 170 175 180  
 Pro Ser Ala Leu Ile Ala Leu Ala Pro Leu Leu Phe Gly Arg Gly  
 185 190 195  
 Glu Val Cys Asp Ala Arg Leu Gln Arg Cys Gln Val Ser Arg Glu  
 200 205 210  
 Pro Ser Tyr Ala Ala Phe Ser Thr Arg Gly Ala Phe His Leu Pro  
 215 220 225  
 Leu Gly Val Val Pro Phe Val Tyr Arg Lys Ile Tyr Glu Ala Ala  
 230 235 240  
 Lys Phe Arg Phe Gly Arg Arg Arg Arg Ala Val Leu Pro Leu Pro  
 245 250 255  
 Ala Thr Met Gln Val Lys Val Lys Glu Ala Pro Asp Glu Ala Glu  
 260 265 270  
 Val Val Phe Thr Ala His Cys Lys Ala Thr Val Ser Phe Gln Val  
 275 280 285  
 Ser Gly Asp Ser Trp Arg Glu Gln Lys Glu Arg Arg Ala Ala Met  
 290 295 300  
 Met Val Gly Ile Leu Ile Gly Val Phe Val Leu Cys Trp Ile Pro  
 305 310 315  
 Phe Phe Leu Thr Glu Leu Ile Ser Pro Leu Cys Ala Cys Ser Leu  
 320 325 330  
 Pro Pro Ile Trp Lys Ser Ile Phe Leu Trp Leu Gly Tyr Ser Asn  
 335 340 345  
 Ser Phe Phe Asn Pro Leu Ile Tyr Thr Ala Phe Asn Lys Asn Tyr  
 350 355 360  
 Asn Asn Ala Phe Lys Ser Leu Phe Thr Lys Gln Arg  
 365 370

<210> 9  
 <211> 330

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474848CD1

&lt;400&gt; 9

Met	Asp	Pro	Thr	Thr	Pro	Ala	Trp	Gly	Thr	Glu	Ser	Thr	Thr	Val
1				5					10					15
Asn	Gly	Asn	Asp	Gln	Ala	Leu	Leu	Leu	Leu	Cys	Gly	Lys	Glu	Thr
				20					25					30
Leu	Ile	Pro	Val	Phe	Leu	Ile	Leu	Phe	Ile	Ala	Leu	Val	Gly	Leu
				35					40					45
Val	Gly	Asn	Gly	Phe	Val	Leu	Trp	Leu	Leu	Gly	Phe	Arg	Met	Arg
				50					55					60
Arg	Asn	Ala	Phe	Ser	Val	Tyr	Val	Leu	Ser	Leu	Ala	Gly	Ala	Asp
				65					70					75
Phe	Leu	Phe	Leu	Cys	Phe	Gln	Ile	Ile	Asn	Cys	Leu	Val	Tyr	Leu
				80					85					90
Ser	Asn	Phe	Phe	Cys	Ser	Ile	Ser	Ile	Asn	Phe	Pro	Ser	Phe	Phe
				95					100					105
Thr	Thr	Val	Met	Thr	Cys	Ala	Tyr	Leu	Ala	Gly	Leu	Ser	Met	Leu
				110					115					120
Ser	Thr	Val	Ser	Thr	Glu	Arg	Cys	Leu	Ser	Val	Leu	Trp	Pro	Ile
				125					130					135
Trp	Tyr	Arg	Cys	Arg	Arg	Pro	Arg	His	Leu	Ser	Ala	Val	Val	Cys
				140					145					150
Val	Leu	Leu	Trp	Ala	Leu	Ser	Leu	Leu	Ser	Ile	Leu	Glu	Gly	
				155					160					165
Lys	Phe	Cys	Gly	Phe	Leu	Phe	Ser	Asp	Gly	Asp	Ser	Gly	Trp	Cys
				170					175					180
Gln	Thr	Phe	Asp	Phe	Ile	Thr	Ala	Ala	Trp	Leu	Ile	Phe	Leu	Phe
				185					190					195
Met	Val	Leu	Cys	Gly	Ser	Ser	Leu	Ala	Leu	Leu	Val	Arg	Ile	Leu
				200					205					210
Cys	Gly	Ser	Arg	Gly	Leu	Pro	Leu	Thr	Arg	Leu	Tyr	Leu	Thr	Ile
				215					220					225
Leu	Leu	Thr	Val	Leu	Val	Phe	Leu	Leu	Cys	Gly	Leu	Pro	Phe	Gly
				230					235					240
Ile	Gln	Trp	Phe	Leu	Ile	Leu	Trp	Ile	Trp	Lys	Asp	Ser	Asp	Val
				245					250					255
Leu	Phe	Cys	His	Ile	His	Pro	Val	Ser	Val	Val	Leu	Ser	Ser	Leu
				260					265					270
Asn	Ser	Ser	Ala	Asn	Pro	Ile	Ile	Tyr	Phe	Phe	Val	Gly	Ser	Phe
				275					280					285
Arg	Lys	Gln	Trp	Arg	Leu	Gln	Gln	Pro	Ile	Leu	Lys	Leu	Ala	Leu
				290					295					300
Gln	Arg	Ala	Leu	Gln	Asp	Ile	Ala	Glu	Val	Asp	His	Ser	Glu	Gly
				305					310					315
Cys	Phe	Arg	Gln	Gly	Thr	Pro	Glu	Met	Ser	Arg	Ser	Ser	Leu	Val
				320					325					330

&lt;210&gt; 10

&lt;211&gt; 494

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7655614CD1

&lt;400&gt; 10-

Met	Glu	Glu	Pro	Gln	Pro	Pro	Arg	Pro	Pro	Ala	Ser	Met	Ala	Leu
1				5					10					15
Leu	Gly	Ser	Gln	His	Ser	Gly	Ala	Pro	Ser	Ala	Ala	Gly	Pro	Pro
			20						25					30
Gly	Gly	Thr	Ser	Ser	Ala	Ala	Thr	Ala	Ala	Val	Leu	Ser	Phe	Ser
			35						40					45
Thr	Val	Ala	Thr	Ala	Ala	Leu	Gly	Asn	Leu	Ser	Asp	Ala	Ser	Gly
			50						55					60
Gly	Gly	Thr	Ala	Ala	Ala	Pro	Gly	Gly	Gly	Gly	Leu	Gly	Gly	Ser
			65						70					75
Gly	Ala	Ala	Arg	Glu	Ala	Gly	Ala	Ala	Val	Arg	Arg	Pro	Leu	Gly
			80						85					90
Pro	Glu	Ala	Ala	Pro	Leu	Leu	Ser	His	Gly	Ala	Ala	Val	Ala	Ala
			95						100					105
Gln	Ala	Leu	Val	Leu	Leu	Leu	Ile	Phe	Leu	Leu	Ser	Ser	Leu	Gly
			110						115					120
Asn	Cys	Ala	Val	Met	Gly	Val	Ile	Val	Lys	His	Arg	Gln	Leu	Arg
			125						130					135
Thr	Val	Thr	Asn	Ala	Phe	Ile	Leu	Ser	Leu	Ser	Leu	Ser	Asp	Leu
			140						145					150
Leu	Thr	Ala	Leu	Leu	Cys	Leu	Pro	Ala	Ala	Phe	Leu	Asp	Leu	Phe
			155						160					165
Thr	Pro	Pro	Gly	Gly	Ser	Ala	Pro	Ala	Ala	Ala	Ala	Gly	Pro	Trp
			170						175					180
Arg	Gly	Phe	Cys	Ala	Ala	Ser	Arg	Phe	Phe	Ser	Ser	Cys	Phe	Gly
			185						190					195
Ile	Val	Ser	Thr	Leu	Ser	Val	Ala	Leu	Ile	Ser	Leu	Asp	Arg	Tyr
			200						205					210
Cys	Ala	Ile	Val	Arg	Pro	Pro	Arg	Glu	Lys	Ile	Gly	Arg	Arg	Arg
			215						220					225
Ala	Leu	Gln	Leu	Leu	Ala	Gly	Ala	Trp	Leu	Thr	Ala	Leu	Gly	Phe
			230						235					240
Ser	Leu	Pro	Trp	Glu	Leu	Leu	Gly	Ala	Pro	Arg	Glu	Leu	Ala	Ala
			245						250					255
Ala	Gln	Ser	Phe	His	Gly	Cys	Leu	Tyr	Arg	Thr	Ser	Pro	Asp	Pro
			260						265					270
Ala	Gln	Leu	Gly	Ala	Ala	Phe	Ser	Val	Gly	Leu	Val	Val	Ala	Cys
			275						280					285
Tyr	Leu	Leu	Pro	Phe	Leu	Leu	Met	Cys	Phe	Cys	His	Tyr	His	Ile
			290						295					300
Cys	Lys	Thr	Val	Arg	Leu	Ser	Asp	Val	Arg	Val	Arg	Pro	Val	Asn
			305						310					315
Thr	Tyr	Ala	Arg	Val	Leu	Arg	Phe	Phe	Ser	Glu	Val	Arg	Thr	Ala
			320						325					330
Thr	Thr	Val	Leu	Ile	Met	Ile	Val	Phe	Val	Ile	Cys	Cys	Trp	Gly
			335						340					345
Pro	Tyr	Cys	Phe	Leu	Val	Leu	Leu	Ala	Ala	Ala	Arg	Gln	Ala	Gln
			350						355					360
Thr	Met	Gln	Ala	Pro	Ser	Leu	Leu	Ser	Val	Val	Ala	Val	Trp	Leu
			365						370					375
Thr	Trp	Ala	Asn	Gly	Ala	Ile	Asn	Pro	Val	Ile	Tyr	Ala	Ile	Arg
			380						385					390
Asn	Pro	Asn	Ile	Ser	Met	Leu	Leu	Gly	Arg	Asn	Arg	Glu	Glu	Gly
			395						400					405
Tyr	Arg	Thr	Arg	Asn	Val	Asp	Ala	Phe	Leu	Pro	Ser	Gln	Gly	Pro
			410						415					420
Gly	Leu	Gln	Ala	Arg	Ser	Arg	Ser	Arg	Leu	Arg	Asn	Arg	Tyr	Ala
			425						430					435
Asn	Arg	Leu	Gly	Ala	Cys	Asn	Arg	Met	Ser	Ser	Ser	Asn	Pro	Ala
			440						445					450
Ser	Gly	Val	Ala	Gly	Asp	Val	Ala	Met	Trp	Ala	Arg	Lys	Asn	Pro
			455						460					465

Val	Val	Leu	Phe	Cys	Arg	Glu	Gly	Pro	Pro	Glu	Pro	Val	Thr	Ala
				470					475					480
Val	Thr	Lys	Gln	Pro	Lys	Ser	Glu	Ala	Gly	Asp	Thr	Ser	Leu	
				485					490					

&lt;210&gt; 11

&lt;211&gt; 532

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6792419CD1

&lt;400&gt; 11

Met	Asn	Lys	Ser	Thr	Cys	Leu	Met	Ala	Ala	Glu	Thr	Pro	Ser	Lys
1				5					10					15
Arg	Trp	Arg	Leu	His	Cys	Leu	Ala	Phe	Ser	Gln	Arg	Phe	Val	Arg
				20					25					30
Ala	Gly	Pro	Ala	Cys	Ser	Ser	Arg	Glu	Ala	Cys	Ser	Ser	Pro	Arg
				35					40					45
Ala	Gly	Trp	Asn	Pro	Ala	Gly	Phe	Arg	Leu	Pro	Gly	Arg	Trp	Ser
				50					55					60
Pro	Phe	Val	Ala	Leu	His	Leu	Val	Cys	Gln	Ile	Arg	Glu	Ala	Leu
				65					70					75
Lys	Leu	Arg	Ser	Gly	His	Arg	Thr	Pro	Ser	Gly	Ala	Gly	Ser	Ser
				80					85					90
Met	Gln	Pro	Pro	Pro	Ser	Leu	Cys	Gly	Arg	Ala	Leu	Val	Ala	Leu
				95					100					105
Val	Leu	Ala	Cys	Gly	Leu	Ser	Arg	Ile	Trp	Gly	Glu	Glu	Arg	Gly
				110					115					120
Phe	Pro	Pro	Asp	Arg	Ala	Thr	Pro	Leu	Leu	Gln	Thr	Ala	Glu	Ile
				125					130					135
Met	Thr	Pro	Pro	Thr	Lys	Thr	Leu	Trp	Pro	Lys	Gly	Ser	Asn	Ala
				140					145					150
Ser	Leu	Ala	Arg	Ser	Leu	Ala	Pro	Ala	Glu	Val	Pro	Lys	Gly	Asp
				155					160					165
Arg	Thr	Ala	Gly	Ser	Pro	Pro	Arg	Thr	Ile	Ser	Pro	Pro	Pro	Cys
				170					175					180
Gln	Gly	Pro	Ile	Glu	Ile	Lys	Glu	Thr	Phe	Lys	Tyr	Ile	Asn	Thr
				185					190					195
Val	Val	Ser	Cys	Leu	Val	Phe	Val	Leu	Gly	Ile	Ile	Gly	Asn	Ser
				200					205					210
Thr	Leu	Leu	Arg	Ile	Ile	Tyr	Lys	Asn	Lys	Cys	Met	Arg	Asn	Gly
				215					220					225
Pro	Asn	Ile	Leu	Ile	Ala	Ser	Leu	Ala	Leu	Gly	Asp	Leu	Leu	His
				230					235					240
Ile	Val	Ile	Asp	Ile	Pro	Ile	Asn	Val	Tyr	Lys	Leu	Leu	Ala	Glu
				245					250					255
Asp	Trp	Pro	Phe	Gly	Ala	Glu	Met	Cys	Lys	Leu	Val	Pro	Phe	Ile
				260					265					270
Gln	Lys	Ala	Ser	Val	Gly	Ile	Thr	Val	Leu	Ser	Leu	Cys	Ala	Leu
				275					280					285
Ser	Ile	Asp	Arg	Tyr	Arg	Ala	Val	Ala	Ser	Trp	Ser	Arg	Ile	Lys
				290					295					300
Gly	Ile	Gly	Val	Pro	Lys	Trp	Thr	Ala	Val	Glu	Ile	Val	Leu	Ile
				305					310					315
Trp	Val	Val	Ser	Val	Val	Leu	Ala	Val	Pro	Glu	Ala	Ile	Gly	Phe
				320					325					330
Asp	Ile	Ile	Thr	Met	Asp	Tyr	Lys	Gly	Ser	Tyr	Leu	Arg	Ile	Cys
				335					340					345
Leu	Leu	His	Pro	Val	Gln	Lys	Thr	Ala	Phe	Met	Gln	Phe	Tyr	Lys
				350					355					360

```

Thr Ala Lys Asp Trp Trp Leu Phe Ser Phe Tyr Phe Cys Leu Pro
365 370 375
Leu Ala Ile Thr Ala Phe Phe Tyr Thr Leu Met Thr Cys Glu Met
380 385 390
Leu Arg Lys Lys Ser Gly Met Gln Ile Ala Leu Asn Asp His Leu
395 400 405
Lys Gln Arg Arg Glu Val Ala Lys Thr Val Phe Cys Leu Val Leu
410 415 420
Val Phe Ala Leu Cys Trp Leu Pro Leu His Leu Ser Arg Ile Leu
425 430 435
Lys Leu Thr Leu Tyr Asn Gln Asn Asp Pro Asn Arg Cys Glu Leu
440 445 450
Leu Ser Phe Leu Leu Val Leu Asp Tyr Ile Gly Ile Asn Met Ala
455 460 465
Ser Leu Asn Ser Cys Ile Asn Pro Ile Ala Leu Tyr Leu Val Ser
470 475 480
Lys Arg Phe Lys Asn Cys Phe Lys Ser Cys Leu Cys Cys Trp Cys
485 490 495
Gln Ser Phe Glu Lys Gln Ser Leu Glu Glu Lys Gln Ser Cys
500 505 510
Leu Lys Phe Lys Ala Asn Asp His Gly Tyr Asp Asn Phe Arg Ser
515 520 525
Ser Asn Lys Tyr Ser Ser Ser
530

```

&lt;210&gt; 12

&lt;211&gt; 485

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474790CD1

&lt;400&gt; 12

```

Met Pro Ile Ser Leu Ala His Gly Ile Ile Arg Ser Thr Val Leu
1 5 10 15
Val Ile Phe Leu Ala Ala Ser Phe Val Gly Asn Ile Val Leu Ala
20 25 30
Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln Val Thr Asn Arg
35 40 45
Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln Ile Ser Leu
50 55 60
Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe Trp Pro
65 70 75
Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His Leu
80 85 90
Phe Ala Phe Ala Ser Val Asn Thr Ile Val Val Val Ser Val Asp
95 100 105
Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met
110 115 120
Thr Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val
125 130 135
Ala Ile Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala
140 145 150
Ala Phe Asp Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala
155 160 165
Ser Pro Ser Tyr Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile
170 175 180
Pro Leu Ile Val Met Ile Ala Cys Tyr Ser Val Val Phe Cys Ala
185 190 195
Ala Arg Arg Gln His Ala Leu Leu Tyr Asn Val Lys Arg His Ser
200 205 210

```

```
<210> 13
<211> 255
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 7474816CD1
```

<400> 13														
Met	Pro	Phe	Ile	Ser	Lys	Leu	Val	Leu	Ala	Ser	Gln	Pro	Thr	Leu
1				5					10					15
Phe	Ser	Phe	Phe	Ser	Ala	Ser	Ser	Pro	Phe	Leu	Leu	Phe	Leu	Asp
				20					25					30
Leu	Arg	Pro	Glu	Arg	Thr	Tyr	Leu	Pro	Val	Cys	His	Val	Ala	Leu
				35					40					45
Ile	His	Met	Val	Val	Leu	Leu	Thr	Met	Val	Phe	Leu	Ser	Pro	Gln
				50					55					60
Leu	Phe	Glu	Ser	Leu	Asn	Phe	Gln	Asn	Asp	Phe	Lys	Tyr	Glu	Ala
				65					70					75
Ser	Phe	Tyr	Leu	Arg	Arg	Val	Ile	Arg	Asp	Leu	Ser	Ile	Cys	Thr
				80					85					90
Thr	Cys	Leu	Leu	Gly	Met	Leu	Gln	Val	Val	Asn	Ile	Ser	Pro	Ser
				95					100					105

Ile	Ser	Trp	Leu	Val	Arg	Phe	Lys	Trp	Lys	Ser	Thr	Ile	Phe	Thr	110	115	120
Phe	His	Leu	Phe	Ser	Trp	Ser	Leu	Ser	Phe	Pro	Val	Ser	Ser	Ser	125	130	135
Leu	Ile	Phe	Tyr	Thr	Val	Ala	Ser	Ser	Asn	Val	Thr	Gln	Ile	Asn	140	145	150
Leu	His	Val	Ser	Lys	Tyr	Cys	Ser	Leu	Phe	Pro	Ile	Asn	Ser	Ile	155	160	165
Ile	Arg	Gly	Leu	Phe	Phe	Thr	Leu	Ser	Leu	Phe	Arg	Asp	Val	Phe	170	175	180
Leu	Lys	Gln	Ile	Met	Leu	Phe	Ser	Ser	Val	Tyr	Met	Met	Thr	Leu	185	190	195
Ile	Gln	Glu	Leu	Gln	Glu	Ile	Leu	Val	Pro	Ser	Gln	Pro	Gln	Pro	200	205	210
Leu	Pro	Lys	Asp	Leu	Cys	Arg	Gly	Lys	Ser	His	Gln	His	Ile	Leu	215	220	225
Leu	Pro	Val	Ser	Phe	Ser	Val	Gly	Met	Tyr	Lys	Met	Asp	Phe	Ile	230	235	240
Ile	Ser	Thr	Ser	Ser	Thr	Leu	Pro	Trp	Ala	Tyr	Asp	Arg	Gly	Val	245	250	255

&lt;210&gt; 14

&lt;211&gt; 881

&lt;212&gt; PRT

&lt;213&gt; Homo\_sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476172CD1

&lt;400&gt; 14

Met	Ser	Ile	Glu	Glu	Leu	Cys	Ser	Asp	Phe	Lys	Lys	Tyr	Leu	Phe	1	5	10	15
Pro	Asn	Ser	Phe	Glu	Ile	Ser	Val	Phe	Leu	Gln	Thr	Leu	Ala	Met	20	25	30	
Ile	His	Ser	Ile	Glu	Met	Ile	Asn	Asn	Ser	Thr	Leu	Leu	Pro	Gly	35	40	45	
Val	Lys	Leu	Gly	Tyr	Glu	Ile	Tyr	Asp	Thr	Cys	Thr	Glu	Val	Thr	50	55	60	
Val	Ala	Met	Ala	Ala	Thr	Leu	Arg	Phe	Leu	Ser	Lys	Phe	Asn	Cys	65	70	75	
Ser	Arg	Glu	Thr	Val	Glu	Phe	Lys	Cys	Asp	Tyr	Ser	Ser	Tyr	Met	80	85	90	
Pro	Arg	Val	Lys	Ala	Val	Ile	Gly	Ser	Gly	Tyr	Ser	Glu	Ile	Thr	95	100	105	
Met	Ala	Val	Ser	Arg	Met	Leu	Asn	Leu	Gln	Leu	Met	Pro	Gln	Val	110	115	120	
Gly	Tyr	Glu	Ser	Thr	Ala	Glu	Ile	Leu	Ser	Asp	Lys	Ile	Arg	Phe	125	130	135	
Pro	Ser	Phe	Leu	Arg	Thr	Val	Pro	Ser	Asp	Phe	His	Gln	Ile	Lys	140	145	150	
Ala	Met	Ala	His	Leu	Ile	Gln	Lys	Ser	Gly	Trp	Asn	Trp	Ile	Gly	155	160	165	
Ile	Ile	Thr	Thr	Asp	Asp	Asp	Tyr	Gly	Arg	Leu	Ala	Leu	Asn	Thr	170	175	180	
Phe	Ile	Ile	Gln	Ala	Glu	Ala	Asn	Asn	Val	Cys	Ile	Ala	Phe	Lys	185	190	195	
Glu	Val	Leu	Pro	Ala	Phe	Leu	Ser	Asp	Asn	Thr	Ile	Glu	Val	Arg	200	205	210	
Ile	Asn	Arg	Thr	Leu	Lys	Lys	Ile	Ile	Leu	Glu	Ala	Gln	Val	Asn	215	220	225	
Val	Ile	Val	Val	Phe	Leu	Arg	Gln	Phe	His	Val	Phe	Asp	Leu	Phe				

Asn Lys Ala Ile	230	Asn Lys Met Trp Ile Ala Ser	235	240
245	Met Asn Ile Asn Lys Met Trp Ile Ala Ser	250	255	
Asp Asn Trp Ser	260	Thr Thr Ile Pro Asn Val	265	270
Lys Lys Ile Gly	275	Lys Val Val Gly Phe Ala Phe Arg Arg Gly Asn	280	285
Ile Ser Ser Phe	290	His Ser Phe Leu Gln Asn Leu His Leu Leu Pro	295	300
Ser Asp Ser His	305	Lys Leu Leu His Glu Tyr Ala Met His Leu Ser	310	315
Ala Cys Ala Tyr	320	Val Lys Asp Thr Asp Leu Ser Gln Cys Ile Phe	325	330
Asn His Ser Gln	335	Arg Thr Leu Ala Tyr Lys Ala Asn Lys Ala Ile	340	345
Glu Arg Asn Phe	350	Val Met Arg Asn Asp Phe Leu Trp Asp Tyr Ala	355	360
Glu Pro Gly Leu	365	Ile His Ser Ile Gln Leu Ala Val Phe Ala Leu	370	375
Gly Tyr Ala Ile	380	Arg Asp Leu Cys Gln Ala Arg Asp Cys Gln Asn	385	390
Pro Asn Ala Phe	395	Gln Pro Trp Glu Leu Leu Gly Val Leu Lys Asn	400	405
Val Thr Phe Thr	410	Asp Gly Trp Asn Ser Phe His Phe Asp Ala His	415	420
Gly Asp Leu Asn	425	Thr Gly Tyr Asp Val Val Leu Trp Lys Glu Ile	430	435
Asn Gly His Met	440	Thr Val Thr Lys Met Ala Glu Tyr Asp Leu Gln	445	450
Asn Asp Val Phe	455	Ile Ile Pro Asp Gln Glu Thr Lys Asn Glu Phe	460	465
Arg Asn Leu Lys	470	Leu Thr Leu Phe Ser Val Leu Thr Lys Leu Lys	475	480
His Gln Lys Arg	485	Ile Pro Val Ala Thr Val Thr Ser Val Pro Val	490	495
Pro Leu Pro Ser	500	Ile Trp His Tyr Arg Gln Thr Val Cys Ala Pro	505	510
Ser Gln Asp Met	515	Pro His Cys Leu Leu Cys Asn Asn Lys Thr His	520	525
Trp Ala Pro Val	530	Arg Ser Thr Met Cys Phe Glu Lys Glu Val Glu	535	540
Tyr Leu Asn Trp	545	Asn Asp Ser Leu Ala Ile Leu Leu Leu Ile Leu	550	555
Ser Leu Leu Gly	560	Ile Ile Phe Val Leu Val Val Gly Ile Ile Phe	565	570
Thr Arg Asn Leu	575	Asn Thr Pro Val Val Lys Ser Ser Gly Gly Leu	580	585
Arg Val Cys Tyr	590	Val Ile Leu Leu Cys His Phe Leu Asn Phe Ala	595	600
Ser Thr Ser Phe	605	Ile Gly Glu Pro Gln Asp Phe Thr Cys Lys	610	615
Thr Arg Gln Thr	620	Met Phe Gly Val Ser Phe Thr Leu Cys Ile Ser	625	630
Cys Ile Leu Thr	635	Lys Ser Leu Lys Ile Leu Leu Ala Phe Ser Phe	640	645
Asp Pro Lys Leu	650	Gln Lys Phe Leu Lys Cys Leu Tyr Arg Pro Ile	655	660
Leu Ile Ile Phe	665	Thr Cys Thr Gly Ile Gln Val Val Ile Cys Thr	670	675
Leu Trp Leu Ile	680	Phe Ala Ala Pro Thr Val Glu Val Asn Val Ser	685	690
Leu Pro Arg Val	695	Ile Ile Leu Glu Cys Glu Glu Gly Ser Ile Leu	700	705



Ala Phe Gly Thr	Met Leu Gly Tyr Ile	Ala Ile Leu Ala Phe Ile	
	710	715	720
Cys Phe Ile Phe	Ala Phe Lys Gly Lys	Tyr Glu Asn Tyr Asn Glu	
	725	730	735
Ala Lys Phe Ile	Thr Phe Gly Met Leu	Ile Tyr Phe Ile Ala Trp	
	740	745	750
Ile Thr Phe Ile	Pro Ile Tyr Ala Thr	Thr Phe Gly Lys Tyr Val	
	755	760	765
Pro Ala Val Glu	Ile Ile Val Ile Leu	Ile Ser Asn Tyr Gly Ile	
	770	775	780
Leu Tyr Cys Thr	Phe Ile Pro Lys Cys	Tyr Val Ile Ile Cys Lys	
	785	790	795
Gln Glu Ile Asn	Thr Lys Ser Ala Phe	Leu Lys Met Ile Tyr Ser	
	800	805	810
Tyr Ser Ser His	Ser Val Ser Ser Ile	Ala Leu Ser Pro Ala Ser	
	815	820	825
Leu Asp Ser Met	Ser Gly Asn Val Thr	Met Thr Asn Pro Ser Ser	
	830	835	840
Ser Gly Lys Ser	Ala Thr Trp Gln Lys	Ser Lys Asp Leu Gln Ala	
	845	850	855
Gln Ala Phe Ala	His Ile Cys Arg Glu	Asn Ala Thr Ser Val Ser	
	860	865	870
Lys Thr Leu Pro	Arg Lys Arg Met Ser	Ser Ile	
	875	880	

&lt;210&gt; 15

&lt;211&gt; 309

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472141CD1

&lt;400&gt; 15

Met Leu Leu Asn His	Thr Leu Ile Thr Glu	Phe Leu Leu Leu Gly	
1	5	10	15
Val Thr Asp Ile Gln	Glu Leu Asn Pro Ile	Leu Phe Val Met Val	
	20	25	30
Leu Ala Met Tyr Phe	Ile Asn Val Phe Gly	Asn Gly Ala Ile Met	
	35	40	45
Met Ile Val Ile Leu	Asp Pro Arg Leu Tyr	Ser Pro Met Tyr Phe	
	50	55	60
Phe Leu Gly Asn Leu	Ala Cys Leu Asp Ile	Cys Phe Ser Thr Val	
	65	70	75
Thr Val Pro Lys Met	Leu Glu Asn Phe Phe	Ser Thr Ser Lys Ala	
	80	85	90
Ile Ser Phe Leu Gly	Cys Ile Thr Gln Leu	His Phe Phe His Phe	
	95	100	105
Leu Gly Ser Thr Glu	Ala Leu Leu Leu Thr	Val Met Ala Phe Asp	
	110	115	120
Arg Phe Val Ala Ile	Cys Arg Pro Leu His	Tyr Pro Val Ile Met	
	125	130	135
Asn Arg Gln Leu Cys	Ile His Met Thr Val	Thr Ile Trp Thr Ile	
	140	145	150
Gly Phe Phe His Ala	Leu Leu His Ser Val	Met Thr Ser Arg Leu	
	155	160	165
Ser Phe Cys Gly Ser	Asn His Ile His His	Phe Phe Cys Asp Val	
	170	175	180
Lys Pro Leu Leu Asp	Leu Ala Cys Gly Asn	Thr Glu Leu Asn Leu	
	185	190	195
Trp Leu Leu Asn Thr	Val Thr Gly Thr Ile	Ala Leu Thr Ser Phe	
	200	205	210

```

Tyr Leu-Ile Phe Leu Ser Tyr Phe Tyr Ile Ile Thr Asn Leu Leu
215 220 225
Leu Lys Thr Arg Ser Cys Ser Met Leu His Lys Ala Leu Ser Thr
230 235 240
Cys Ala Ser His Phe Met Val Val Val Leu Phe Tyr Ala Pro Val
245 250 255
Leu Phe Thr Tyr Ile Arg Pro Ala Ser Gly Ser Ser Leu Asp Gln
260 265 270
Asp Thr Ile Ile Ala Ile Met Tyr Ser Val Val Thr Pro Ala Leu
275 280 285
Asn Pro Leu Met Tyr Thr Leu Arg Asn Lys Glu Val Arg Ser Ala
290 295 300
Leu Asn Arg Lys Val Arg Ser Ser Leu
305

```

```

<210> 16
<211> 224
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472137CD1

```

```

<400> 16
Met Arg Asn Phe Ser Val Val Ser Glu Phe Ile Leu Leu Gly Ile
1 5 10 15
Pro His Thr Glu Gly Leu Glu Thr Ile Leu Leu Val Leu Phe Leu
20 25 30
Ser Phe Tyr Ile Phe Thr Leu Met Gly Asn Leu Leu Ile Leu Leu
35 40 45
Ala Ile Val Ser Ser Ala Arg Leu His Thr Pro Met Tyr Phe Phe
50 55 60
Leu Cys Lys Leu Ser Val Phe Asp Leu Phe Phe Pro Ser Val Ser
65 70 75
Ser Pro Lys Met Leu Cys Tyr Leu Ser Gly Asn Ser Arg Ala Ile
80 85 90
Ser Tyr Ala Gly Cys Ala Ser Gln Leu Phe Phe Tyr His Phe Leu
95 100 105
Gly Cys Thr Glu Cys Phe Leu Tyr Thr Val Met Ala Tyr Asp Arg
110 115 120
Phe Val Ala Ile Cys His Pro Leu Arg Tyr Thr Ile Ile Met Ser
125 130 135
His Arg Ala Cys Ile Ile Leu Ala Met Gly Thr Ser Phe Phe Gly
140 145 150
Cys Ile Gln Ala Thr Phe Leu Thr Thr Leu Thr Phe Gln Leu Pro
155 160 165
Tyr Cys Val Pro Asn Glu Val Asp Tyr Tyr Phe Cys Asp Ile Pro
170 175 180
Val Met Leu Lys Leu Ala Cys Ala Asp Thr Ser Ala Leu Glu Met
185 190 195
Val Gly Phe Ile Ser Val Gly Leu Met Pro Leu Ser Cys Phe Leu
200 205 210
Leu Ile Leu Thr Ser Tyr Ser Gly Ile Val Phe Ser Ile Leu
215 220

```

```

<210> 17
<211> 326
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature

```

&lt;223&gt; Incyte ID No: 7477934CD1

&lt;400&gt; 17

```

Met Gly His Gln Asn His Thr Phe Ser Ser Asp Phe Ile Leu Leu
 1          5          10          15
Gly Leu Phe Ser Ser Asn Lys Cys Gly Leu Leu Leu Arg Gln Phe
          20          25          30
Val Ile Phe Ile Met Ser Val Thr Glu Asn Thr Leu Met Ile Leu
          35          40          45
Leu Ile Arg Ser Asp Ser Arg Leu His Thr Pro Met Tyr Phe Leu
          50          55          60
Leu Ser His Leu Ser Leu Met Asp Ile Leu His Val Ser Asn Ile
          65          70          75
Val Pro Lys Met Val Thr Asn Phe Leu Ser Gly Ser Arg Thr Ile
          80          85          90
Ser Phe Ala Gly Cys Gly Phe Gln Val Phe Leu Ser Leu Thr Leu
          95          100          105
Leu Gly Gly Glu Cys Leu Leu Leu Ala Ala Met Ser Cys Asp Arg
          110          115          120
Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Pro Ile Leu Met Lys
          125          130          135
Glu Tyr Ala Ser Ala Leu Met Ala Gly Gly Ser Trp Leu Ile Gly
          140          145          150
Val Phe Asn Ser Thr Val His Thr Ala Tyr Ala Leu Gln Phe Pro
          155          160          165
Phe Cys Gly Ser Arg Ala Ile Asp His Phe Phe Cys Glu Val Pro
          170          175          180
Ala Met Leu Lys Leu Ser Cys Ala Asp Thr Thr Arg Tyr Glu Arg
          185          190          195
Gly Val Cys Val Ser Ala Val Ile Phe Leu Leu Ile Pro Phe Ser
          200          205          210
Leu Ile Ser Ala Ser Tyr Gly Gln Ile Ile Leu Thr Val Leu Gln
          215          220          225
Met Lys Ser Ser Glu Ala Arg Lys Lys Ser Phe Ser Thr Cys Ser
          230          235          240
Phe His Met Ile Val Val Thr Met Tyr Tyr Gly Pro Phe Ile Phe
          245          250          255
Thr Tyr Met Arg Pro Lys Ser Tyr His Thr Pro Gly Gln Asp Lys
          260          265          270
Phe Leu Ala Ile Phe Tyr Thr Ile Leu Thr Pro Thr Leu Asn Pro
          275          280          285
Phe Ile Tyr Ser Phe Arg Asn Lys Asp Val Leu Ala Val Met Lys
          290          295          300
Asn Met Leu Lys Ser Asn Phe Leu His Lys Lys Met Asn Arg Lys
          305          310          315
Ile Pro Glu Cys Val Phe Cys Leu Phe Leu Cys
          320          325

```

&lt;210&gt; 18

&lt;211&gt; 2374

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1714538CB1

&lt;400&gt; 18

```

gtcagtcagt ccactggctc ccgcgcgcgc tctgtgtccg tcgctcggag ggtggaagcc 60
ggggtctcgc gggccgcggg ccgcatgact cctctctgcc tcaattgctc tgcctccct 120
ggagacctgt acccaggggg tgcaagggaac cccatggctt gcaatggcag tgccggccagg 180
gggcactttg accctgagga cttgaacctg actgacgagg cactgagact caagtacctg 240
gggccccagc agacagagct gttcatgccc atctgtgcca catacctgct gatcttcgtg 300

```

```

gtgggctgctg tgggcaatgg gctgacctgt ctggtcatcc tgcgccacaa ggccatgcgc 360
acgcctacca actactacct cttcagcctg gccgtgtcgg acctgctggt gctgctggtg 420
ggcctgcccc tggagctcta tgagatgtgg cacaactacc ccttcctgct gggcggttgg 480
ggctgtatt tccgcacgct actgtttgag atggtctgcc tggcctcagt gctcaacgtc 540
actgccctga gcgtggaacg ctatgtggcc gtggtgcacc cactccaggc caggtccatg 600
gtgacgcggg cccatgtgcg ccgagtgtt ggggcgtct ggggtcttgc catgctctgc 660
tccctgcca acaccagcct gcacggcatc cggcagctgc acgtgccctg ccggggccca 720
gtgccagact cagctgtttg catgctggtc cgccacggg ccctctacaa catggtagt 780
cagaccaccg cgctgtctt cttctgcctg cccatggcca tcatgagcgt gctctgcctg 840
ctcgttgggc tgcactgcg gcgaggagg ctgctgctca tgcaggaggc caagggcagg 900
ggctctgcag cagccaggtc cagatacacc tgcaggctcc agcagcacga tcggggccgg 960
ggacaagtga ccaagatgct gtttgcctg gtcgtggtgt ttggcatctg ctggggcccg 1020
ttccacgccg accgcgtcat gtggagcgtc gtgtcacagt ggacagatgg cctgcacctg 1080
gccttcacgc acgtgcacgt catctccggc atctctctt acctgggctc ggcggccaac 1140
cccgctctct atagcctcat gtccagccgc ttccgagaga cctccaggga ggccctgtgc 1200
ctcggggcct gctgccatcg cctcagacc cgcacagct cccacagcct cagcaggatg 1260
accacaggca gcaccctgtg tgatgtgggc tccctgggca gctgggtcca cccctggct 1320
gggaacgatg gcccagaggc gcagcaagag accgatccat cctgagtggg gccttaaagt 1380
ggcttcacct gggaggccca gagggtcacc tggagctggg gagacacatc tgccttcctc 1440
tgcagggatg ccttcacgta ctgtccctag ttccagcctg aaattctgac cagcacctca 1500
gtttccctca gagggaaca gcaggaggag ggatccctga ctgctgagga ctcacactga 1560
ccagacgcca caccttgtgc ttcttatctg tccactgcca ctccccaggt tcaaatcctt 1620
accctgcaga aatatcacag ttagctggg ctcagcagtc ctccctctgg ggaactccctg 1680
ccaccactgc cagtttctga aacgggtcca ctgggtcctc actgtccttc ccagttcctg 1740
ttcaggttct ggcaggggcc cagggatcca ggggacctg ttccaatctc agccctgctg 1800
tcaccacctt gtcctgcacc atcaagcata tcagtctacc tttcttttt tctgagacag 1860
agtctcactc tgtcgcacc gctggagtgc agtggcgcca ttttgactcg ctgcaacctc 1920
cgctccggg gttcaggcga ttctcctgcc tcagcctccc gactagctgg gactacaggt 1980
gagccccagc atgcccagct aattttttt agtttttagt agagacgggg tttcaccatg 2040
ttggccaggc tgggtctcga ctcttgacct caggtgatcc gccgacctcg gcctcccaaa 2100
gtcctcggtat tacaggcatg agccaccaca cccggccaat cagtccacct ttctaggcct 2160
tggttccttg cctgaaaaat gaaagaggcg ctggctttcc acagtgtcat gctttggcac 2220
tttagctatg gttttcttct tgtgtgtgtg taagccactg cttataataa aaccaacaat 2280
accctcagac tgaagggcg gaagttatta tctgatctt tatcaacccc aagccccact 2340
tctcctctga cctccccatg cctccccag cctc 2374

```

&lt;210&gt; 19

&lt;211&gt; 813

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3406743CB1

&lt;400&gt; 19

```

ctcttcgggtg cctcttcttc ctccgggaca aggatggagg atctcttttag cccctcaatt 60
ctgccgcggg cgcccaacat ttccgtgccc atcttctggt gctggggtct caacctgacc 120
ttggggcaag gagccctgc ctctgggccc cccagccgcc gcgtccgcct ggtgttctg 180
ggggtcatcc tgggtgtggc ggtggcaggc aacaccacag tgctgtgccg cctgtgcggc 240
ggcgggcggg cctgggcccgg ccccaagcgt cgcaagatgg acttctctgct ggtgcagctg 300
gccctggcgg acctgtacgc gtgcgggggc acggcgctgt cacagctggc ctgggaactg 360
ctgggagagc cccgcgcggc cacgggggac ctggcgtgcc gcttcctgca gctgtgcag 420
gcatccgggc ggggcgcctc ggcccacctc gtggtgctca tgcacctega gcgcggcgc 480
gcgccaggcg cgccactctc cgcccgagcc tggccgggga tgcgtcgtg cactggatc 540
ttcgcgctcc tgcagcgtg gcacgtgcag gtgtacgct tctataggc cgtcgcgggc 600
ttcgtcgcgc ctgtaagat catgggtgtc gcttgtggcc acctactctc cgtctggtgg 660
cggcaccggc tgaaggcccc agcgggtgca gggcctggt cggcgagccc aggtggagcc 720
cgtgcgccc cgcgatgcc ccgcgccaag gtgcagagcc tgaagatgag ccagctgctg 780
gggtgctgt tcgtgggctg cgagctgccc tag 813

```

&lt;210&gt; 20

&lt;211&gt; 1542

<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3485895CB1

<400> 20  
tgaataaatt attcctgggc ttttgaatta ttaatcttgg ctgcaggata tgcaaatagg 60  
atatgcaaatt aggacacttg gcacaacact gggtcagaag ctatatccag ctgctggcag 120  
agttcctgtc aagggatcaa gtcttccaac agaattggta tgggtttaact cagcagaatt 180  
tgttgaacaa ctacgacatg ctgggggatca tggcatggaa tgcaacttgc aaaaactggc 240  
tggcagcaga ggctgccctg gaaaagtact acctttccat tttttatggg attgagttcg 300  
ttgtgggagt ccttgggaaat accattgttg ttacggcta catcttctct ctgaagaact 360  
ggaacagcag taatatattat ctctttaacc tctctgtctc tgacttagct tttctgtgca 420  
ccctcccat gctgataagg agttatgcca atggaaactg gatatatgga gacgtgctct 480  
gcataagcaa ccgatatgtg ctcatgcca acctctatac cagcattctc tttctcactt 540  
ttatcagcat agatcgatac ttgataatta agtatccttt ccgagaacac cttctgcaaa 600  
agaaagagtt tgctatttta atctccttgg ccatttgggt tttagtaacc ttagagttac 660  
taccataact tccccttata aatcctgtta taactgacaa tggcaccacc tgtaattgatt 720  
ttgcaagttc tggagacccc aactacaacc tcatttacag catgtgtcta acactgttgg 780  
ggttccttat tctctttttt gtgatgtgtt tcttttatta caagattgct ctcttcctaa 840  
agcagaggaa taggcaggtt gctactgtct tggcccttga aaagcctctc aacttgggtca 900  
tcatggcagt gctaattctt tctgtgcttt ttacacccta tcacgtcatg cggaatgtga 960  
ggatcgcttc agcctgggg agttggaagc agtatcagt cactcaggtc gtcataact 1020  
ccttttacat tgtgacacgg cctttggcct ttctgaacag tgtcatcaac cctgtcttct 1080  
attttctttt gggagatcac ttcagggaca tgctgatgaa tcaactgaga cacaacttca 1140  
aatcccttac atccttttagc agatgggctc atgaactcct actttcattc agagaaaagt 1200  
gaggggcttg tgaacagat tgttctacag atgaatctgt aagccagtta cagtttgctc 1260  
taactcatag acatcaatca gagagtgtca cagatttaac cttgatctaa agacaagttg 1320  
taccagagt atgtgaaaag aatgggacga caagaatgta ctgggttctt cctctaagaa 1380  
ttgaaaggag ttgaactgcc ttatgtttgg gcattgtaact ccaaaatact aggtagtata 1440  
aggctttctc aatcagtgca aaaatggaag atatataaag caacaagttg tctgcatttg 1500  
atcactgggc agattgtaaa aaaaaaaaaa aagggcggcc gc 1542

<210> 21  
<211> 1191  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7476102CB1

<400> 21  
atgggggatg agctggcacc ttgccctgtg ggcactacag cttggccggc cctgatccag 60  
ctcatcagca agacaccctg catgccccaa gcagccagca acacttcctt gggcctggg 120  
gacctcaggg tgcccagctc catgctgtac tggcttttcc ttccctcaag cctgctggct 180  
gcagccacac tggctgtcag cccctgtctg ctggtgacca tctgcgga ccaacggctg 240  
cgacaggagc cccactacct gctcccggct aacatcctgc tctcagacct ggcctacatt 300  
ctctccaca tgctcatctc ctccagcagc ctgggtggct gggagctggg ccgcatggcc 360  
tgtggcattc tactgatgc tgtcttcgcc gcctgcacca gcaccatcct gtccttcacc 420  
gccattgtgc tgcaaccta cctggcagtc atccatccac tgcgtacct ctcttcattg 480  
tcccattggg ctgcctggaa ggcagtggcc ctcatctggc tgggtggcctg ctgcttccc 540  
acattcctta ttgggtcag caagtggcag gatgccagc tggaggagca aggagttca 600  
tacatcctac caccaagcat gggcaccag cgggatgtg gcctcctggc cattgttacc 660  
tacacctcca ttctgtcgt tctgttctc tgcaagctc tcattgcca ctgtttctgg 720  
aggatctatg cagaggccaa gacttcaggc atctgggggc agggctattc cggggccagg 780  
ggcaccctgc tgatccactc agtctgtac acattgtac tgagcacagg ggtggtgtc 840  
tccttgga caagtgtgac caggtaccac cacattgact ctgggactca cacatggctc 900  
ctggcagcta acagtgaggt actcatgat ctctcccgtg ccatgctcac atacctgtac 960  
ctgctccgct accggcagct gttgggcatg gtccggggcc acctccatc caggaggcac 1020  
caggccatct ttaccatttc ctgctgtctg gaaccactat tctacttttc tctgattttg 1080

accactctag gaggatgat tattccattg tgtgtagaga ccacattttg tttatccatt 1140  
catctgtcaa tggaccgttt ggggtgcacc acctttggct attgtgaata a 1191

<210> 22

<211> 3360

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2432942CB1

<400> 22

cccacgcgtc cggcagctga gacggcagcg gcagcttctc agggccggag ccagttcttg 60  
gaggagactc tgcgcagggc atggatcact gtggtgccct tttcctgtgc ctgtgccttc 120  
tgactttgca gaatgcaaca acagagacat gggagaact cctgagctac atggagaata 180  
tgacaggtgtc cagggggccg agctcagttt tttcctctcg tcaactccac cagctggagc 240  
agatgtact gaacaccagc ttcccaggct acaacctgac cttgcagaca cccaccatcc 300  
agtctctggc cttcaagctg agctgtgact tctctggcct ctcgctgacc agtgccactc 360  
tgaagcgggt gcccagga gagggtcagc atgcccggg tcagcacgcc atgcagttcc 420  
ccgcccagct gaccggggac gcctgcaaga cccgccccag ggagctgcgg ctcatctgta 480  
tctacttctc caacacccac tttttcaagg atgaaaacaa ctcatctctg ctgaataact 540  
acgtcctggg gggccagctg agtcatgggc acgtgaacaa cctcagggat cctgtgaaca 600  
tcagcttctg gcacaacaa agcctggaag gctacacct gacctgtgtc tctggaagg 660  
agggagccag gaaacagccc tgggggggct ggagccctga gggctgtcgt acagagcagc 720  
cctccactc tcaggtgtc tgccgtgca accacctcac ctactttgct gttctcatgc 780  
aactctcccc agcctgggtc cctgcagagt tctgtggcacc tcttacgtac atctccctcg 840  
tgggctgcag catctccatc gtggcctcgc tgatcacagt cctgtgtcac ttccatttca 900  
ggaagcagag tgactcctta acacgcaccc acatgaacct gcatgcctcc gtgtgctcc 960  
tgaacatcgc tttcctgtcg agccccgcat tcgcaatgtc tctgtgtccc gggtcagcat 1020  
gcacggctct gggcgtgccc ctgactacg cgctgtcag ctgcctcacc tggatggcca 1080  
tcgagggtct caacctctac ctctcctcg ggcgtgtcta caacatctac atccgcagat 1140  
atgtgttcaa gcttggtgtg ctaggctggg gggccccagc cctcctgggt ctgctttccc 1200  
tctctgtcaa gagctcggta tacggacct gcacaatccc cgtcttcgac agctgggaga 1260  
atggcacagc cttccagaac atgtccatat gctgggtgcg gagccccgtg gtgcacagtg 1320  
tctgtgtcat gggctacggc ggctcacgt cctcttcaa cctgggtggg ctggcctggg 1380  
cgctgtggac cctgcgcagg ctgcccggagc gggcggatgc accaagtgtc agggcctgcc 1440  
atgacactgt cactgtgtcg ggccctaccg tgctgtgtgg aaccacctgg gccttggcct 1500  
tcttttctt tggcgtcttc ctgtgcccc agctgttctt cttcaccatc ttaacctcgc 1560  
tgtacggttt cttccttttc ctgtggttct gctcccagc gtgcccgtca gaagcagagg 1620  
ccaaggcaca gatagaggcc ttcagctcct cccaaacaac acagtagtcc gggcctcctg 1680  
gcctggaatc ctacgcctct ctggccgcca gtagcctgag gctacggctc ctgctagaga 1740  
gggtggcagg cctgtgtgtg gaccccagag gccactgtga ccgccaaggg gccttttcca 1800  
cttccacggc ctctccaggc actgagggga aggcattgct ctacctctcc ctgacatttt 1860  
gctccggggc agatccaacc ttacctgggg cagcaaacct tgtcctggta cctgggcca 1920  
gctcgcagg gatgtgggca gagcaccagc ctgggcatca ggaagccaag tttcaaggac 1980  
tgtctttgag tctgtctgta tgacctggg cctgccactt ctacagacc ctaggatatc 2040  
acagctgtga catgggggca agcagctttg tttcagccta acccaggagc ttagtaaaaa 2100  
ttgcataaga ccagggggaa gagtgtcagc gtgggggtgg aattcccgcg gcctccacct 2160  
gcttgctagg ggcaggatct cattcaggct gccctggaag cacctgcttg gccctgccac 2220  
cttccctcag gggagggcca gatggcatcc tggcttgggg cgggtgggac ctaccaggc 2280  
tctgagactt tactggccta tgctgaggc ctcttttctt ttaactccct aaattatgat 2340  
gactccaagt ccaagcccac ccttcccaaa gattgggagg ttccgcgctt cccagaggct 2400  
cctcctggcg tgctcccaag acttccatag accatctgga ccagtagccc atcccgcagt 2460  
tttcttgggg ccagaggaaa acgcttcttt ctctccagc tgaatcagct ggtatccagt 2520  
gtcctggctg tttggtgatt gggcaagatt gaatttgccc aggtaggcgt gagagtgtgg 2580  
gttttaaat cgaagctcag gccatagttt cagagaatca cccttaccac agaccttcat 2640  
gagacagtgc tcatgaagcc agtgcgtttc ccagaacgaa cactaggcgg caccgttgg 2700  
ccacactcag aggccttgg cgccaagact gcatctagaa tcgctcaaac acctgtttgc 2760  
agaccccatg caccagctgg aggggcccga atgcaggac tggcctact gagtaccaca 2820  
tttccctcag gaggaaaggc aagacacgct tacacggcca tttgtctctt tcccaatgc 2880  
ggcgggtcac tttcgtctt gggggctgca cccagacat agctggcacc agagcagggt 2940  
gctcagggtg tgggtgctca gggccctgcc ccaggccact gggccgtttt gatgacctg 3000

```

aagggtcacāg gcagaaaaata ggagcaggat ttccccctggg gaaaagtctt cctgggacat 3060
cttctgtctct tctgtacatt tctagatgca aataactcct tcaccaggca gtgagtggcg 3120
taggctctggg agccaggctg cctgggctcc aatgccagct ctgccacttg ctagctgtga 3180
gactgtgggac aaaccactca gcctctgtgt gcctcagttt tcctatttgt aaaatagagg 3240
ccatagtggg acctattttg aagactaagt aaaagaattc aaataaagag acttggcaca 3300
gagtaagtgc tcagtaaaaa aaaaaaaggg ggggcgccga aaaaaggctc cgaaccggaa 3360

```

&lt;210&gt; 23

&lt;211&gt; 1660

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4630911CB1

&lt;400&gt; 23

```

gtagatgaac aggaagcaac tggaggtaat acctatgatt tgggtaatat ctctagataa 60
gagtaaaggg acttttcaca taggcagtca cttcacctct acgagtctca ttttttccaa 120
ctgcacaaat acagattttc gatactttat ttatgcagtg acatacactg tcattcttgt 180
gccagggtctc atagggaata tattagccct gtgggtattc tatggttata tgaaagaaac 240
aaaacgagct gtgatattta tgataaactt agccattgct gacttactac aagttctttc 300
cttgccactg aggatcttct actacttgaa tcatgactgg ccatttgggc ctggtctctg 360
catgttctgt ttctacctga agtatgtcaa catgtatgca agcatctact tcttgggtctg 420
catcagtgtg cgacgatttt ggtttctcat gtaccccttt cgcttccatg actgcaaaca 480
gaaatatgac ctgtacatca gcattgctgg ctggctgac atctgccttg cctgtgtact 540
ctttccactc ctcagaacca gtgatgatac ccctggcaat aggaccaaat gctttgtgga 600
tcttcttacc aggaatgtca acctggccca gtccgttgtt atgatgacca ttggcgagtt 660
gattgggttt gtaactccgc ttctgattgt cctatatgtt acctggaaga cggttttatc 720
actgcaagat aaatatccca tggcccaaga tcttggagag aaacagaaag ccttgaagat 780
gattctaacc tgtgcagggg tattccta attgctttgca ccttatcatt tcagttttcc 840
tttagatttc ctggtgaagt ccaatgaaat taaaagctgc ctagccagaa ggggtgattct 900
aatatttcat tctgtggcat tgtgtcttgc tagtctgaat tcatgtcttg acccagtcac 960
atactacttt tccactaatg agttccgaag acggctttca agacaagatt tgcātggacag 1020
catccaactc catgcaaaat cctttgtgag taaccataca gcttccacca tgacacctga 1080
attatgctaa aacaaaaaac caaactgaat gtgacctgaa atgcaagtac atcagaacat 1140
atctgcaata cccaagccac agggaagaac ttgcaaaaca acacagcttt tcagtctctgc 1200
tctatcttac tgctatgggg aattcacttc ttcaaagcag gacctatttg gagcattacg 1260
atccacgatt attgatgttg acatgtccat gtagtaattt ttcttcaagt ctgtaaatct 1320
taaaatatca aatttctgtg acatcctata aacatatgca ctcaactgta gtcagtactt 1380
ttacctgtga acctcaggca caaaaagatt attagcttgg agtcaccata aacatttagt 1440
tttggtgcaa cagatactga gtctttatgt tcagaggaaa tgtaagtgtc tttttatatt 1500
agtaatcaca gtttacatgc cattttcata tgtttggtta tatttttagtg ggatagatga 1560
tatattacc tgtgtaaaaga aaatgtaaac ataagatcat ttttatctct caagtgtgat 1620
tactcttcag agtttgaaga ttaaaatagc tattttcttc 1660

```

&lt;210&gt; 24

&lt;211&gt; 2603

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472432CB1

&lt;400&gt; 24

```

ggaccagcag aaggaagaag tatggagtta aagactgcag cgtgaactga ggagtcccgg 60
acaggccgct tgctgcagag gatccagtc agatcccagg agagcccctc tgccccttcg 120
gacctcgtct cccatctaca aaacgtgaag attggcccag tttagcgtgtc tctacaaaaa 180
gggtcatata cactgcagcc gctgcaggct gatctgagaa agcctctggc ccaccgccat 240
ggccttctct atgcacctgc tggctctgct cttcggaatg ggctcctggg tgaccatcaa 300
tgggctctgg gtagagctgc ccctgctggg gatggagctg cccgagggct ggtacctgcc 360
ctcctacctc acggtgggtca tccagctggc caacatcggg cccctcctgg tcacctgtct 420

```

```

ccatcacttc cggcccagct gcctttccga agtgcccatg atcttcaccc tgctgggctg 480
gggaaccgct acctgcatca tctttgcctt cctctggaat atgacctcct ggggtgctga 540
cggccaccac agcatcgccct tcttggtcct cacttcttct ctggccctgg tggactgcac 600
ctcttcagtg accttctctg cgttcatgag cgggtgccc acctactacc tcaccacctt 660
ctttgtgggt gaaggactca gcggcctctt gccgcctctg gtggctcttg cccagggtctc 720
cgggtctcact acctgcgtca atgtcactga gatatcagac agcgtaccaaa gccctgtacc 780
cacgagggag actgacatcg cacagggagt tcccagagct ttggtgtccg ccctccccgg 840
aatggaagca cccttgctcc acctggagag ccgctacctt cccgccact tctcacccct 900
ggtcttcttc ctctcctat ccatcatgat ggctgtctgc ctctggtcgt tctttgtcct 960
ccagcgtcaa cccagggtgct gggaggcttc cgtggaagac ctctcaatg accagggtcac 1020
cctccactcc atccggccgc ggggaagagaa tgacttgggc cctgcaggca cgggtggacag 1080
cagccagggc caggggtatc tagaggagaa agcagccccc tgctgcccgg cgcacctggc 1140
cttcatctat accctgggtg ccttcgtcaa cgcgtcacc aacggcatgc tgccctctgt 1200
gcagacctac tcttgccctg agttgcctac cacttggtg ccacctcag 1260
cattgtggcc aacctctctg cctcgttgggt ctccatgttc ctgcctaaca ggtctctgct 1320
gttcttggg gtctctccg tgcctgggac ctgctttggg ggctacaaca tggccatggc 1380
ggtgatgagc ccctgcccc tcttgccagg ccactggggt ggggaagtcc tcattgtggc 1440
ctcgtgggtg cttttcagcg gctgcctcag ttacgtcaag gtgatgtgg gcgtggctct 1500
gcgcgacctc agccgcagcg cctcttctgt gtgcggggcg gcggtgcagc tgggctcgt 1560
gctcggagcg ctgctcatgt tccctctggt caacgtgctg cggctcttct cgtcccgga 1620
cttctgcaat ctgcactgtc cagcctaggg aggcgcgcca ccccgcccc atcgctcacg 1680
gacggaactg ggggtccagag aggcaggtc acagagcaag gggcaggaa agagagacag 1740
agcctgagta attgaatcat gaacgcaagt gcccactggg gactgtgggg aagatggcac 1800
ctggaaatgc aaggtgcggc tctatcccca actctgtgtc acactacctg tgacgaccag 1860
ctcagatctc ctttgccttg actctcaaga gaggactgat ttgcagcatc tagctggagg 1920
caggcccaag ggtgttagaa gggaaacagc tgggacagcc ggctgtccct tcaggctgtg 1980
tgaccttggg aaagtcatct ggcttctctg tgccctgttc ttcatgcatg cagtgggat 2040
tccagtaagt accaactacc tcacaggcat ggcacggagg caaaaggaaa aagcagcccc 2100
catcaagcaa gccctcctgg gccacctgct gatctgacag tccatcgtag taacaagagt 2160
ggcagtctgc acaacctaga agtgccaga aggggtgaga cagccccctg ccctctctcc 2220
tttgcctctc agtctcacag aggggtctct acaagacaag cagataacga tagaatcttg 2280
ggcatcttgg ctttcggatt ctcatgtgtg agggacgtag taccacacac acccttctct 2340
gtcatccttc ctggcccata aagcccacta gttggagagt aagtaccctc ctggaagcag 2400
ggagagatga tttgctggtg gggctgggga agcccctcc ctgagcctct gaaagtgaac 2460
tccccgacca ggttggggac cagacatgca gagcccctgg aagtattctc tcaaatggag 2520
gcaacagagg tgattgttat tttgttttag tttctgtttt tcattttttt aaataaaggc 2580
attccctgct tttaaaaaaa aaa 2603

```

&lt;210&gt; 25

&lt;211&gt; 1119

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474977CB1

&lt;400&gt; 25

```

atggaggcgc ctaccctttc agtgccacc gccggcggtg cccttgccct gggacccgag 60
accagcagcg ggaccccaag cccgagagg atactcggtt cgaccccgag cggcgccgtc 120
ctgcccggcc gagggccgcc cttctctgtc ttacagggtc tgggtggtgac gctgctagtg 180
ctgctgatcg ccgccacttt cctgtggaac ctgctgggtc cggtcacat cccgcgggtc 240
cgtgccttcc accgcgtgcc gcataacttg gtggcctcga cggcctctc ggacgaacta 300
gtggcagcgc tggcgatgcc accgagcctg gcgagtgagc tgtcgaccgg gcgacgtcgg 360
ctgctgggccc ggagcctgtg ccacgtgtgg atctccttcc acgtgctgtg ctgccccgcc 420
ggcctcgagg acgtggcgcc catcgccctg ggccgcgacg gggccatcac acggcacctg 480
cagcacacgc tgcgcacccg cagccgcgcc tcgttgctca tgatcgcgct caccgggtg 540
ccgtcgggcg tcatcgccct cgcgcgcgtg ctctttggcc ggggcgaggt gtgcgacgct 600
cggctccagc gctgccaggt gagccgggaa ccctcctatg ccgcttctc caccgcggc 660
gccttccacc tgcgccttgg cgtgggtgcc tttgtctacc ggaagatcta cgaggcgcc 720
aagtttctgt tcggccgcgc ccggagagct gtgctgccgt tgccggccac catgcaggtg 780
aaggtaaagg aagcacctga tgaggctgaa gtggtgttca cggcacattg caaagcaacg 840
gtgtccttcc aggtgagcgg ggactcctgg cgggagcaga aggagaggcg agcagccatg 900

```



```

atgggtgggaa ttctgattgg cgtgtttgtg ctgtgctgga tccccctctt cctgacggaa 960
ctcatcagcc cactctgtgc ctgcagcctg cccccatct ggaaaagcat atttctgtgg 1020
cttggctact ccaattcttt cttcaacccc ctgatttaca cagcttttaa caagaactac 1080
aacaatgcct tcaagagcct ctttactaag cagagatga 1119

```

&lt;210&gt; 26

&lt;211&gt; 1018

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474848CB1

&lt;400&gt; 26

```

gggcaccagt ggaggttttc tgagcatgga tccaaccacc cgggcctggg gaacagaaag 60
tacaacagtg aatggaaatg accaagccct tcttctgctt tgtggcaagg agaccctgat 120
cccgtctctc ctgatacctt tcattgccct ggtcgggctg gtaggaaacg ggtttgtgct 180
ctggctcctg ggcttccgca tgcgcaggaa cgccttctct gtctacgtcc tcagcctggc 240
cggggccgac ttccctcttcc tctgcttcca gattataaat tgcctgggtg acctcagtaa 300
cttcttctgt tccatctcca tcaatttccc tagcttcttc accactgtga tgacctgtgc 360
ctaccttgca ggcctgagca tgctgagcac cgtcagcacc gagcgtgcc tgtccgtcct 420
gtggcccatc tggatatcgt gccgcgcgcc cagacacctg tcagcggctg tgtgtgtcct 480
gctctggggc ctgtccctac tgcctgagcat cttggaaggg aagttctgtg gcttcttatt 540
tagtgatggt gactctgggt ggtgtcagac atttgatttc atcactgcag cgtggctgat 600
ttttttattc atggttctct gtgggtccag tctggccctg ctggtcagga tctctgtggt 660
ctccaggggt ctgccactga ccaggtgta cctgaccatc ctgctcacag tgcctggtgt 720
cctcctctgc ggcctgccct ttggcattca gtggttccta atattatgga tctggaagga 780
ttctgatgtc ttattttgtc atattcatcc agtttcagtt gtctgtcat ctcttaacag 840
cagtgcacac cccatcattt acttcttctg gggtctttt aggaagcagt ggcggtgca 900
gcagccgcat ctcaagctgg ctctccagag ggctctgcag gacattgctg aggtggatca 960
cagtgaagga tgcttccgtc agggcaccac ggagatgtcg agaagcagtc tgggtgatg 1018

```

&lt;210&gt; 27

&lt;211&gt; 2177

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7655614CB1

&lt;400&gt; 27

```

atggaggagc cgcagccgcc ccgcccacca gcgagcatgg ccttactggg cagccagcac 60
tccggcgccc cctcccgccc cggcccacct ggcgggactt cctcccgccc caccggcgcc 120
gtgctctcct tcagcacctg ggcgaccgcg gcgctgggga acctgagcga cgcaagcgga 180
ggcggcacag ctgccgctcc cggtgggcgg gcccttggcg ggtccggggc agcgcgggag 240
gccccggcgg cggtgaggcg gccgctaggg ccggaggcgg cgcgctgct gtcgcacgga 300
gctgcagtgg cggcccaggc gctcgtctct ctgctcatct tctgctgtc tagccttggc 360
aactgcgcgg tgatgggggt gattgtgaag caccggcagc tccgcaccgt caccaacgcc 420
ttcatcctgt cgtgtccct atcggatctg ctacaggcgc tgctctgctt gcccgcgcc 480
ttcctggacc tcttcactcc gcccgggggt tcggcgccct cgcgcgccgc ggggcccctg 540
cgcggcttct gcgcgccag ccgcttcttc agctcgtgct tcggcatcgt gtccacgctc 600
agcgtggcgc tcactctggt ggaccgttac tgcgtatcgt tgcggccgcc gcgggagaag 660
atcggccgce gccgcgcgt cagctgctg gcgggcgccc ggtgacggc cctgctcttc 720
tccttgccct gggagctgct gggggcgcgc cgggaactcg cggcgccgca gagcttccac 780
ggctgcctct accggacctc cccggacccc gcgcagctgg gcgcggcctt cagcgtgggg 840
ctggtgggtg cctgctacct gctgcccttc ctgctcatgt gcttctgcca ctaccacatc 900
tgcaagacgg tgcgcctgtc ggacgtgcgc gtgcggccgg tgaacacctc cgcgcgcgtg 960
ctgcgcttct tcagcgaggt gcgcacggcc accacgctcc tcactcatgat cgtcttctgc 1020
atctgctgct tggggcccta ctgcttctg gtgctgctgg ccgcgcgccg gcaggcccag 1080
accatgcagg cccctcgtct cctcagcgtg gtggccgtct ggctgacctg ggccaatggg 1140
gccatcaacc ctgtcatcta cgccatccgc aatcccaaca tttcgatgct cctagggcgc 1200

```

```

aaccgcgagg agggctaccg gactaggaat gtggacgctt tcctgcccag ccagggccccg 1260
ggtctgcaag ccagaagccg cagtcgcctt cgaaaccgct atgccaaccg gctggggggcc 1320
tgcaacagga tgtcctcttc caaccgggcc agcggagtg gggggagcgt ggccatgtgg 1380
gcccgcacaa atccagtgtg acttttctgc cgagagggac caccagagcc ggtgacggca 1440
gtgaccaaac agcctaaatc cgaagctggg gataccagcc tctaagacgg ttggaatggc 1500
cagcttatga aggcaaattt ccactcgc attttaatga tgggaagattc tgggggagag 1560
ttgtggattt cataaagcca aacattttaa gctagagacg ggggaggctt accactttcc 1620
ccaaacaaca taaaagacaa tgtcccttct tcaaaagtgc caaaaggaa gtataatgca 1680
aaaattaaaa caatcttaaa ccacataacc aagcattgtg aactgtaagt gccaaaaatg 1740
acaaaaataa cattcactat aactgaaaag ctcataatat aggaccacac tgtgaaacaa 1800
acaaaacatt gaatgcaacc agtattgttc aactacacag aatttcagga atgaatggag 1860
accttcagag ctgcttgaaa gccctcttaa atcgctacca tccaacaaaa ctttagcttc 1920
tagagcttgc acctctgaca tgctttgggt gcttagattg caagtggaaa agtcttatcc 1980
ttacacacat tgggtgcacct cctcacccca cccctaata acctttgtgg aaaattttta 2040
attcataacc catttaaaat catttttaga ttttggaaag ctggttatta ttagataaaa 2100
ttcaaccatc tgaaggacaa atagaagtat gaacataact ttccagcaca ctgcgccgta 2160
ctagtgtcca ggctcgt 2177

```

&lt;210&gt; 28

&lt;211&gt; 1632

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6792419CBI

&lt;400&gt; 28

```

ttgatgggaa gggatgaatg aataaaaagta cttgtctgat ggcagcagag accccgagca 60
aacggtggag gctacactgt ctggcattct cgcagcgttt cgtcagagcc ggacccgctt 120
gcagctcaag ggaggcgtgc tcctctccca gagcaggctg gaaccagcgt gggttccgcc 180
tcccgggaag gtggtctcca ttcgtcgtc tgcactgtgt ttgtcagatc cgagaggctc 240
tgaaactgcg gagcggccac cggacgcctt ctggagcagg tagcagcatg cagccgcctc 300
caagtctgtg cggacgcgcc ctggttgccg tggttcttgc ctgcggcctg tcgcggatct 360
ggggagagga gagaggcttc ccgcctgaca gggccactcc gcttttgcaa accgcagaga 420
taatgacgcc acccactaag accttatggc ccaagggttc caacgccagt ctggcgcggt 480
cgttggcacc tgcggaggtg cctaaaggag acaggacggc aggatctccg ccacgcacca 540
tctcccctcc ccctgccaag ggacccatcg agatcaagga gactttcaaa tacatcaaca 600
cggttgtgtc ctgccttggt ttcgtgctgg ggatcatcgg gaactccaca cttctgagaa 660
ttatctacaa gaacaagtgc atgcgaaacg gtcccaatat cttgatcgcc agcttggctc 720
tgggagacct gctgcacatc gtcattgaca tccctatcaa tgtctacaag ctgctggcag 780
aggactggcc atttgagct gagatgtgta agctggtgcc tttcatacag aaagcctccg 840
tgggaatcac tgtgctgagt ctatgtgctc tgagtattga cagatcga gctgtgtgctt 900
cttgagtag aattaaagga attggggttc caaaatggac agcagtagaa attgttttga 960
tttgggtggt ctctgtggtt ctggctgtcc ctgaagccat aggttttgat ataattacga 1020
tggactacaa aggaagttat ctgcgaatct cgttgcttca tcccgttcag aagacagctt 1080
tcatgcagtt ttacaagaca gcaaaagatt ggtggctgtt cagtttctat ttctgcttgc 1140
cattggccat cactgcattt ttttatacac taatgacctg tgaaatgttg agaaagaaaa 1200
gtggcatgca gattgcttta aatgatcacc taaagcagag acgggaagtg gccaaaaccg 1260
tcttttgctt ggtccttgct tttgcccctt gctggcttcc ccttcacctc agcaggattc 1320
tgaagctcac tctttataat cagaatgatc ccaatagatg tgaacttttg agctttctgt 1380
tggatttggg ctatattggt atcaacatgg cttcactgaa ttcctgcatt aaccaattg 1440
ctctgtatct ggtgagcaaa agattcaaaa actgctttaa gtcattgctta tgctgtggt 1500
gccagtcatt tgaagaaaaa cagtccttgg aggaaaaagca gtcgtgctta aagttcaaa 1560
ctaattgatc cggatatgac aacttccgtt ccagtaataa atacagctca tcttgaaaga 1620
agaaaaaaa aa 1632

```

&lt;210&gt; 29

&lt;211&gt; 1458

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474790CB1

&lt;400&gt; 29

```
atgcccatca gcctggccca cggcatcatc cgtcaaccg tgctggttat cttcctcgcc 60
gcctctttcg tcggcaacat agtgtggcg ctagtggtgc agcgcaagcc gcagctgctg 120
caggtgacca accgttttat ctttaacctc ctcgtcaccg acctgctgca gatttgcgctc 180
gtggcccccct ggggtggtggc cacctctgtg cctctcttct ggccctcaa cagccacttc 240
tgcacggccc tggtagcct caccacctg ttcgccttcg ccagcgtaa caccattgtc 300
gtggtgtcag tggatcgcta cttgtccatc atccaccctc tctcctaccc gtccaagatg 360
accagcgccc gcggttacct gctcctctat ggcacctgga ttgtggccat cctgcagagc 420
actcctccac tctacggctg gggccaggct gcctttgatg agcgcaatgc tctctgctcc 480
atgatctggg gggccagccc cagctacact attctcagcg tgggtgcctt catcgtcatt 540
ccactgattg tcatgattgc ctgctactcc gtggtgttct gtgcagcccg gaggcagcat 600
gctctgctgt acaatgtcaa gagacacagc ttggaagtgc gagtcaagga ctgtgtggag 660
aatgaggatg aagagggagc agagaagaag gaggagtcc aggatgagag tgagtctcgc 720
cgccagcatg aaggtgaggt caaggccaag gagggcagaa tggaaagcaa ggacggcagc 780
ctgaaggcca aggaaggaaag cacggggacc agtgagagta gtgtagaggc caggggcagc 840
gaggaggtca gagagagcag cacggtggcc agcgacggca gcatggaggg taaggaagggc 900
agcaccaaag ttgaggagaa cagcatgaag gcagacaagg gtcgcacaga ggtcaaccag 960
tgcagcattg acttgggtga agatgacatg gagtttggtg aagacgacat caatttcagt 1020
gaggatgacg tcgaggcagt gaacatcccg gagagcctcc caccagtcg tcgtaacagc 1080
aacagcaacc ctcctctgcc caggtgtac cagtgcaggc ctgctaaagt gatcttcac 1140
atcattttct cctatgtgct atccctgggg ccctactgct ttttagcagt cctggccggtg 1200
tgggtggatg tcgaaaccca ggtaccccag tgggtgatca ccataatcat ctggcttttc 1260
ttcctgcagt gctgcattca cccctatgtc ttgggtaca tgcacaagac cattaagaag 1320
gaaatccagg acatgctgaa gaagttcttc tgcaaggaaa agccccgaa agaagatagc 1380
cacccagacc tgcccggaac agagggtggg actgaaggca agattgtccc ttcctacgat 1440
tctgctactt ttccttga 1458
```

&lt;210&gt; 30

&lt;211&gt; 1015

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474816CB1

&lt;400&gt; 30

```
tgatatatat agtgatatata tacatacatg cacacatata ctatatatat caaggatttta 60
tatgatagga ttaattaaga aaaaaattag tgggaataaaa ataatgttta tgataatttt 120
ggccatagaa tatataatac agatgatgtg aagtacaaaa tgtttttttat acttcatatt 180
ttgatgtaca aagtatgttt gtctttgtaa ttcagatgat tactttgcac ttgtgttccc 240
atgaaaaatg cctttcattt ctaagctggg attggcatct cagccaacac ttttctcctt 300
cttttctgcg tcttctcctt ttctgctttt tctgcatctc aggccagagc gcacttacct 360
accagtctgt catgtggccc tcatccacat ggtggctcct ctcaccatgg tgttcttgtc 420
tccacagctc tttgaatcac tgaattttca gaatgacttc aaatatgagg catctttcta 480
cctgaggagg gtgatcaggg acctctccat ttgtaccacc tgcctcctgg gcatgctgca 540
ggtcgtcaac atcagcccca gcatttctct gttgggtgagg tttaaatgga aatccacaat 600
ttttaccttc catttggtct catggtctct cagtttctct gttagtagta gcctgatctt 660
ttacactgtg gcttcttcca atgtgacca gatcaatttg catgtcagta aatactgttc 720
acttttccca ataaactcca taatcagagg actgttttct actctgtcat tattcagaga 780
tgtttttctt aaacaaataa tgctgttctc aagtgtctac atgatgactc tcattcagga 840
actacaggag atcctggtac cttcacagcc ccagcctcta cctaaggatc ttgacagagg 900
caagagccat cagcacatcc tgctgccggt ggtggcatgt acaagatgga 960
cttcatcatc tcaacctcct caacgttgcc atgggcatat gaccgtggtg tctag 1015
```

&lt;210&gt; 31

&lt;211&gt; 2781

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476172CB1

&lt;400&gt; 31

```
atggcaattct taattatact aattacctgc tttgtgatta ttcttgctac ttcacagcct 60
tgccagacccc ctgatgactt tgtggctgcc acttctccgg gacatatcat aattggaggt 120
ttgtttgcta ttcataaaaa aatgtttgtcc tcagaagact ctcccagacg accacaaatc 180
caggagtgtg ttggccttga aatatcagtt tttcttcaaa ctcttgccat gatacacagc 240
attgagatga tcaacaattc aacactctta cctggagtca aactggggta tgaaatctat 300
gacacttgta cagaagtca agtggaatg gcagccactc tgaggtttct ttctaaattc 360
aactgctcca gagaaactgt ggagtttaag tgtgactatt ccagctacat gccaaagagt 420
aaggctgtca taggttctgg gtactcagaa ataactatgg ctgtctccag gatgttgaat 480
ttacagctca tgccacaggt gggttatgaa tcaactgcag aaatcctgag tgacaaaatt 540
cgctttcctt catttttacg gactgtgccc agtgacttcc atcaaattaa agcaatggct 600
cactgtattc agaaatctgg ttggaactgg attggcatca taaccacaga tgatgactat 660
ggacgattgg ctcttaacac ttttataaatt caggctgaag caaataacgt gtgcatagcc 720
ttcaaagagg ttcttcacgc cttcttttca gataatacca ttgaagttag aatcaatcgg 780
acactgaaga aaatcatttt agaagcccag gttaatgtca ttgtgggtatt tctgaggcaa 840
ttccatgttt ttgatctctt caataaagcc attgaaatga atataaataa gatgtggatt 900
gctagttaga attggtcaac tgccaccaag attaccacca ttcctaattg taaaaagatt 960
ggcaaagttg tagggtttgc ctttagaaga gggaatatat cctctttcca ttcctttctt 1020
caaaatctgc acttgcttcc cagtgcacgt cacaactctt tacatgaata tgccatgcat 1080
ttatctgcct gcgcataatgt caaggacact gatttgagtc aatgcatatt caatcattct 1140
caaaggactt tggcctacaa ggctaacaag gctatagaaa ggaacttcgt catgagaaa 1200
gacttctctt gggactatgc tgagccagga ctcatccta gtattcagct tgcagtgttt 1260
gcccttggtt atgccattcg ggatctgtgt caagctctgt actgtcagaa cccaacgcc 1320
tttcaaccat gggagtact tgggtgtgcta aaaaatgtga cattcactga tggatggaat 1380
tcatttcatt ttgatgtcga cggggattta aatactggat atgatgttgt gctctggaag 1440
gagatcaatg gacacatgac tgtcactaag atggcagaat atgacctaca gaatgatgtc 1500
ttcatcatcc cagatcagga aacaaaaaat gatttcagga atcttaagca aattcaatct 1560
aaatgctcca aggaatgcag tcctgggcaa atgaagaaaa ctacaagaag tcaacacatc 1620
tgttgctatg aatgtcagaa ctgtcctgaa aatcattaca ctaatcagac agatatgcct 1680
cactgccttt tatgcaacaa caaaactcac tgggcccctg ttaggagcac tatgtgcttt 1740
gaaaaggaag tggaatatct caactggaat gactccttgg ccctcctact cctgattctc 1800
tcctacttgg gaatcatatt tgttctggtt gttggcataa tatttacaag aaacctgaac 1860
acacctgttg tgaaatcatc cgggggatta agagtctgct atgtgatcct tctctgtcat 1920
ttctcaatt ttgccagcac gagctttttc attggagaac cacaagactt cacatgtaa 1980
accaggcaga caatgttttg agtgagcttt actctttgca tctcctgcat tttgacgaag 2040
tctctgaaaa ttttgctagc cttcagcttt gatcccaa at tacagaaatt tctgaagtgc 2100
ctctatagac cgatccttat tatcttcaat tgcacgggca tccaggttgt catttgaca 2160
ctctggctaa tctttgcagc acctactgta gaggtgaatg tctccttgcc cagagtcac 2220
atcctggagt gtgaggagg atccatactt gcatttggca ccattgctgg ctacattgcc 2280
atcctggcct tcatttgcct catatttgc ttc aaaggca aatatgagaa ttacaatgaa 2340
gccaaattca ttacatttgg catgctcatt tacttcatag cttggatcac attcaccct 2400
atctatgcta ccacatttgg caaatatgta ccagctgtgg agattattgt catattaata 2460
tctaactatg gaatcctgta ttgcacattc atcccaaat gctatgttat tatttgtaag 2520
caagagatta acacaaagtc tgcctttctc aagatgatct acagttattc tcccatagt 2580
gtgagcagca ttgccctgag tcctgcttca ctggactcca tgagcggcaa tgtcacaatg 2640
accaatccca gctctagtgg caagtctgca acctggcaga aaagcaaaga tcttcaggca 2700
caagcatttg cacacatgt cagggaatat gccacaagt tatctaaac tttgcctcga 2760
aaaagaatgt caagtatatg a
```

&lt;210&gt; 32

&lt;211&gt; 1267

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472141CB1

&lt;400&gt; 32

```

ccccctcct tccccccgga gggggggaaa accatgtaaa ctctcacagc tctaaaaaat 60
ccctttatgt cccttggctt gcatctccta tgaaaagacg ttaataatag aaacgatccc 120
caatttacaa tgcttgaatc gcactacaca ccacaaattc acgcttaata tttttatggg 180
ctgataaggca aatatctaaa acattcacaa gaacttcgcc cgcttttaga aactactctc 240
atgaactatc tttcagctgg catcaactta ttcgcctact aagacttatg gacaagactg 300
ttagtatcc acaaactatc ctatcttttc tttttaaatg ttattgaatc acacattaat 360
cactgaatct ctctctctgg ggggtgacaga catacaagaa ttaaacccta ttctttttgt 420
catggctcctt gccatgtact ttataaatgt gtttgggaat ggagccatca tgatgattgt 480
catcttagat ccaagactct actcacctat gtatttcttc ctgggaaacc tagcatgcct 540
agatatctgc ttctccactg taacagtgcc aaaaatgctg gagaacttct tctctacaag 600
caaagcaatt tccttcttgg gatgcataac tcaacttcat ttcttccact tcttgggaag 660
cacagaagcc ctgctgctga cagtgtggc atttgaccgc tttgtggcta tctgcagacc 720
actccactac cctgtcatca tgaatcgcca gctctgtatt cacatgactg ttactatctg 780
gaccattggc tttttccatg ccctgcttca ctctgtaatg acatctcggt tgagcttctg 840
tgggtccaat catatccatc atttcttctg tgatgttaag ccattactgg atctggcctg 900
tggaacact gagctcaacc tttggctgct caatacagtc acaggcacca ttgccctcac 960
ttccttttat ctgatatttc tctcttattt ctacatcatc accaatcttc tcctcaagac 1020
ccgttcttgc agcatgctcc acaaagctct gtctacctgt gcctcccact tcatggttgt 1080
tggtctgttc tatgctcctg ttctcttcac ctacatccgt cctgcctcag gcagctctct 1140
ggatcaggat acaatcatcg ccatcatgta cagtgtggtc acccctgctc tcaatccact 1200
catgtacacc ttgagaaaca aggaagtggag gagtgcattg aatagaaagg tgagaagctc 1260
cctttga
1267

```

&lt;210&gt; 33

&lt;211&gt; 1559

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472137CB1

&lt;400&gt; 33

```

atagtttgcc tgacatttta gcagatactt cataaaacttt actataatgc taggatattg 60
tgttaccgat gtgttataat aagtatgatt tccattggca gatgatgaaa ctaagtctat 120
gaaagacgtc atgtaaatgt tttaacaagt atagctaata aatggagagg ccaggattca 180
tattaagttt tctggtttct aacctaatla tttttctggt acattgcaga gggcaatgaa 240
cagtggaaat ttcttatgta aagacatcac aagagttaa atagtaata tgctcaaaaa 300
gaataaatcc atggtacata ggcttctaac ttaagtaaaa gtgttttaaa tatgctcta 360
aacttctcat cctgtagaag acaagaatct gcctttacaa agcatccctt ccatgagtcg 420
aatagactga ctctaaaata ggcatttccct tattatctga ttttttttta catggatctc 480
tctcttgccct ttcataagaa ggcaggttgt tccttggtta aaacatagcc aaagcactca 540
tcagcaaaaa gcttgagaac ttgacattgg tggttagact tggtgtcttg ggtacctttt 600
caaggttggg tgttttcaca gccatgagga atttctcggg ggtgtccgaa ttcactctgc 660
tgggcatccc tcacacggag ggtctggaga ctattctggt ggtcctggtt ttgtccttct 720
acatcttcac ccttatgggg aacctgctca tcttctgggc tattgtctcc tctgctcggc 780
ttcacacgcc catgtacttc ttctgtgca agctgtctgt ttttgacctt tttttccctt 840
ctgtgagttc cctaagatg ctgtgctatc tttcaggga cagccgagcc atctcctatg 900
caggctgtgc atcccagctc ttcttctacc atttctcggg ctgcactgag tgtttccctgt 960
acacgggtgat ggcctacgac cgctttgttg ccatttgtca cctctacgc tacaccataa 1020
tcatgagcca cagagcatgt atcatcctag ccatggggac ctcatctctt ggctgcattc 1080
aggccacctt tctgaccact ctcaccttcc aattgcctta ctgtgtcccc aatgaggtgg 1140
actattatct ctgtgatatc ccagtcatgc tgaagctggc ttgtgcagat acctcagccc 1200
tgagatgggt ggggttcac agtggtgggc tcatgcccct cagctgtttc cttctcatcc 1260
tcacctccta cagtggcatc gtcttctcca tctttagat ctgctctgcc gagggccgac 1320
gccgtgcctt ctccacctgc agcgcccacc tcaccgccat cctgcttttt tacatgccag 1380
tggctctcat ttacctgagg cctaccacca gcctgtggtt ggatgcaact gttcaaatc 1440
tgaataacct ggtcaccccc atgctgaacc ccttaatcta cagtctcagg aataaggagg 1500
tgaaattatc actaaggaag gtcttatatc agctgggctt ccttctctgag cagttgtag 1559

```

&lt;210&gt; 34

&lt;211&gt; 981

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477934CB1

&lt;400&gt; 34

```
atggggccatc agaatcacac tttcagcagt gatttcatac ttttgggatt gttctcttct 60
aacaagtgtg gtcttcttct tagacaattt gtcattttca ttatgagtgt aacagaaaaat 120
acgctcatga tcctcctcat tcgcagtgac tcccgaactc acactccaat gtattttctg 180
ctcagccatc tctccttaat ggatactctg catgtttcca acatcgttcc caaaatgggc 240
actaactttc tgtcaggcag cagaactatt tcatttgcag gttgtgggtt ccagggtatt 300
ctgtccctca cctcctctgg tgggtgagtgc cttctcctgg ctgcaatgtc ctgtgatcgc 360
tatgtggcta tctgtcacc gctgcgctat ccgattctta tgaaggagta tgccagcgct 420
ctcatggctg gaggctcctg gctcattggg gttttcaact ccacagtcca cacagcttat 480
gcactgcagt ttcccttctg tggtcttagg gcaattgata acttcttctg tgaagtcctt 540
gcatgttga agttgtcctg tgcagacaca acacgctatg aacgaggggt ttgtgtaagt 600
gctgtgatct tcctgctgat ccttttctcc ttgatctctg cttcttatgg ccaaattatt 660
cttactgtcc tccagatgaa atcatcagag gcaaggaaaa agtcattttc cacttgttcc 720
ttccacatga ttgtggtcac gatgtactat gggccattta tttttacata tatgagacct 780
aaatcatacc acactccagg ccaggataag ttcctggcaa tattctatac gatcctcaca 840
cccacactca accctttcat ctacagcttt aggaataaag atgttctggc ggtgatgaaa 900
aatatgctca aaagtaactt totgcacaaa aaaatgaata ggaaaattcc tgaatgtgtg 960
ttctgtctat ttctatgtta a 981
```

&lt;210&gt; 35

&lt;211&gt; 2177

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7655614CB1.comp

&lt;400&gt; 35

```
acgagcctgg acactagtag ggcgcagtgt gctggaaaagt tatgttcata cttctatttg 60
tccttcagat ggttgaattt tatctaataa taaccagctt ttcaaaatcc taaaatgatt 120
ttaaatgggt tatgaattaa aaattttcca caaagggtat taggggggtg ggtgaggagg 180
tgcaccaatg tgtgtaagga taagactttt ccacttgcaa tctaagcacc caaagcatgt 240
cagaggtgca agctctagaa gctaaagtgt tggtagtgat tagcgattta gaggggcttt 300
caagcagctc tgaagggtct cattcattcc tgaaattctg tgtagttgaa caatactggg 360
tgcattcaat gttttgtttg tttcacagtg tggctctata atatgagctt ttcagttata 420
gtgaatgtta tttttgtcat ttttggcact tacagttcac aatgcttggg tatgtggttt 480
aagattgttt taatttttgc attttacatt ccttttggca cttttgaaga agggacattg 540
tcttttatgt tgtttgggga aagtggtaag cctccccgt ctctagcttt aaatgttttg 600
ctttatgaaa tccacaactc tccccagaa tcttccatca ttaaataatg cgagtggaaa 660
tttgccctca taagctggcc attccaaccg tcttagagggc tggatatccc agcttcggat 720
ttaggtgtt tggctcactgc cgtcaccggc tctggtggtc cctctcggca gaaaagtaca 780
actggatttt tgcggggcca catggccacg tcccctgcca ctccgctggc cgggttgga 840
gaggacatcc tgttgcaggc cccagccgg tttggcatagc ggtttcgaag gcgactgcgg 900
cttctggctt gcagaccgg gccctggctg ggcaggaaaag cgtccacatt cctagtccgg 960
tagccctcct cgcggttgcg ccctaggagg atcgaaatgt tgggatttgc gatggcgtag 1020
atgacagggt tgatggcccc attggcccag gtcagccaga cggccaccac gctgaggagc 1080
gagggggcct gcatggtctg ggctgcccg gcgccggcca gcagcaccag gaagcagtag 1140
ggccccagc agcagatgac gaagacgata atgatgagga cggtggtggc cgtgcgcacc 1200
tcgctgaaga agcgagcac gcgcgcgtag gtgttcaccg gccgcacgcy cactgccagc 1260
aggcgacccg tcttgcagat gtggtagtgg cagaagcaca tgagcaggaa gggcagcagg 1320
tagcaggcca ccaccagccc cagctgaag gccgcgcca gctgcgcggg gtccggggag 1380
gtccggtaga ggcagccgtg gaagctctgc gccgcgcga gttcccgggg cgccccagc 1440
agctcccagg gcaaggagaa gccaggggc gtcagccagg cggccgcccag cagctgcagc 1500
gcgcggcgcc ggcgatctt ctcccggcg gccgcagca tagcgagta acggtccaac 1560
gagatgagcg ccacgctgag cgtggacacg atgccgaagc acgagctgaa gaagcggctg 1620
gcggcgcgca agccgcgcca gggccccgcg gcggcgcgag gcgccaacc cccggggcga 1680
```

gtgaagaggt	ccaggaaggc	ggcgggcagg	cagagcagcg	ccgtgagcag	atccgatagg	1740
gacagcgaca	ggatgaaggc	gttggtgacg	gtgcggagct	gccggtgctt	cacaatcacc	1800
cccatcaccg	cgcagttgcc	aaggctagac	agcaggaaga	tgagcaggag	gacgagcgcc	1860
tgggccgcca	ctgcagctcc	gtgcgacagc	agcggcgccg	cctccggggc	tagcggccgc	1920
ctcaccgccg	ccccgcctc	ccgcgctgcc	ccggacccgc	caaggccgcc	gccaccggga	1980
gcggcagctg	tgccgcctcc	gcttgcgctc	ctcagggttc	ccagcgccgc	ggtcgccacg	2040
gtgctgaagg	agagcacggc	cgccgtggcc	gcggaggaag	tcccgccagg	tgggccggcc	2100
gcggaggggg	cgccggagtg	ctggctgccc	agtaaggcca	tgctcgctgg	tgggcggggc	2160
ggctgcggct	cctccat					2177